REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank) 2001 FEB 12	2. REPORT DATE April 29, 1999	3. REPORT TYPE AND DATES COVERED FINAL 9/30/95-1/29/99

4. TITLE AND SUBTITLE
Multi-Dimensional, Multi-Resolution Adaptive
Processing in the BMDO Setting

5. FUNDING NUMBERS
DAAH04-95-1-0650

6. AUTHOR(S)
Carlos R. Handy, Ph.D.
Romain Murenzi, Ph.D.
Mark J. Smith, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Clark Atlanta University
223 James P. Brawley Drive, SW
Atlanta, GA 30314

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ARO 34933.1-MA-
SD1

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The project aimed at developing new algorithms for tracking missiles and warheads. The initial theoretical foundation for this effort is multiresolution analysis, in particular, continuous multi-dimensional wavelet theory. In concert with this framework, new classes of motion-model techniques for target tracking were developed.

20010222 007

RECEIVED
FEB 13 2001
By

14. SUBJECT TERMS

15. NUMBER OF PAGES
41

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

FINAL REPORT ON CONTRACT DAAH-95-1-0650

Multi-Dimensional, Multi-Resolution Adaptive Processing in the BMDO Setting

Dr. Carlos R. Handy, Project Director
Dr. Romain Murenzi, Co-PI
Dr. Mark J. T. Smith, Co-PI

Center for Theoretical Studies of Physical Systems
Clark Atlanta University
223 James P. Brawley Dr.
Atlanta, Georgia 30314

handyman@ctsps.cau.edu
rmurenzi@ctsps.cau.edu
mjts@eedsp.gatech.edu

1 Introduction

Many techniques have been considered previously for target discrimination and tracking: Kalman-filter techniques, model-based schemes, neural network approaches, classical pattern matching and matched filter methods, dyadic wavelet and subband methods, and so on. All have had limited success in terms of achieving fully reliable automated discrimination. When targets and objects are in motion in an ideal noise-free environment, many techniques will work well. However, under more realistic conditions, these algorithms encounter difficulties. Tracking an object when many other objects are present in the field can be challenging. This situation can occur when a ballistic missile has fragmented. The target, which in this case is the warhead, could be surrounded by tens or hundreds of fragments of similar size. Further complicating the problem are object trajectories that cross and thus introduce periods of occlusion. This can cause tracking algorithms to derail. In addition, the image environment in which these algorithms must work is often noisy. Image acquired from an on-board optical camera located in an interceptor seeker head is typically noisy. Noise has been the "kiss of death" for conventional motion models that have been developed in the context of video coding and motion analysis, such as block matching, optical flow, and pel-recursive algorithms. Finally, useful algorithms must be able to cope with rapidly changing motion and short response-time constraints.

This project is aimed at developing new algorithms for tracking missiles and warheads. The initial theoretical foundation for this effort is multiresolution analysis, in particular, continuous multidimensional wavelet theory. In concert with this framework, we are aggressively developing new classes of motion-model techniques for target tracking, in support of this project.

Thus far, we have pursued several unique approaches.

First we have developed the theory for an algorithm based on the spatio-temporal wavelet transform using Galilean wavelets. This work is described in detail in the paper "Spatio-Temporal Wavelet Transforms for Motion Tracking." It will appear in the Proceedings of ICASSP97 in May, and is included in Appendix A.

Second we have develped a parallel spatio-temporal wavelet approach that uses kinematical wavelets. The transform coefficients are different in this case, resulting in a different complexity/performance tradeoff. Details of this approach are given in the paper "Spatio-Temporal Continuous Wavelets Applied to Missile Warhead Detection and Tracking." This paper will be presented at the SPIE VCIP Conference in February. A copy is included in Appendix B.

Third we have developed a new affine-motion model for estimating motion and used it to develop a model-based method for noise suppression to

facilitate target tracking. Thus far, this algorithm seems to work very well under a broad set of conditions. It is described in the paper "Sensor Data Enhancement of Ballistic Missile Warheads Using an Affine Motion Model." A copy of this paper is included in Appendix C.

Fourth we are combining some of the advances we made with the noise suppression algorithm (appendix C) with the wavelet tracking approach. A discussion of this work in progress is given in Appendix D.

Fifth we have examined a new approach to motion estimation which employs the explicit use of linear prediction. The linear prediction algorithm was developed in a Ph.D. thesis by Robbie Armitano. A discussion of the experimental evaluation is given in Appendix E. The technique works well under mild noise conditions. It has significant advantages over block-based approaches in this regard. However, under the moderate and severe noise conditions we used in our test sequences, the algorithm was not successfull in tracking the motion. Nonetheless, we found this investigation interesting. It could be very useful for sensors with mild noise because it works under these conditions and the complexity is comparable to that of conventional block matching algorithms.

Sixth we examined a simple approach at the beginning of the year, which involved using simple filtering and a Fourier feature representation. This work, although small in scope, gives us some insight into the performance complexity tradeoff. It is very simple from a complexity perspective, yet its results are reasonable. We are presently doing some modifications and comparisons. A discussion of this effort will be given in the final project report, after comparative testing has been completed.

2   Technical Contributors to Project

Faculty:
Romain Murenzi, Associate Professor
Mark J. T. Smith, Professor
Jean-Pierre Leduc, Post-doctoral researcher

Students:
Fernando Mujica
Joseph Monaco
Arthur Redfern
Robbie Armitano
Alen Docef

3

Appendix A

SPATIO-TEMPORAL WAVELET TRANSFORMS FOR MOTION TRACKING. *

Jean-Pierre Leduc   Fernando Mujica   Romain Murenzi   Mark Smith

Georgia Institute of Technology
Center of Signal and Image Processing
Clark Atlanta University
Center for Theoretical Studies of Physical Systems
Atlanta, Georgia

ABSTRACT

This paper addresses the problem of detecting and tracking moving objects in digital image sequences. The main goal is to detect and select mobile objects in a scene, construct the trajectories, and eventually reconstruct the target objects or their signatures. It is assumed that the image sequences are acquired from imaging sensors. The method is based on spatio-temporal continuous wavelet transforms, discretized for digital signal analysis. It turns out that the wavelet transform can be used efficiently in a Kalman filtering framework to perform detection and tracking. Several families of wavelets are considered for motion analysis according to the specific spatio-temporal transformation. Their construction is based on mechanical parameters describing uniform motion, translation, rotation, acceleration, and deformation. The main idea is that each kind of motion generates a specific signal transformation, which is analyzed by a suitable family of continuous wavelets. The analysis is therefore associated with a set of operators that describe the signal transformations at hand. These operators are then associated with a set of selectivity criteria. This leads to a set of filters that are tuned to the moving objects of interest.

1. INTRODUCTION

The primary purpose of the present work is to investigate families of spatio-temporal continuous wavelet transforms (CWT), and to investigate their utility for motion tracking and trajectory constructions. The approach considered in this paper differs fundamentally from other techniques that have been proposed such as those based on optical flow, pel-recursive, block matching and Bayesian models. The main novelty of this method is it combines the CWT with Kalman filtering for tracking. Several families of CWTs can efficiently perform various tasks like motion-based detection and segmentation, selective tracking and reconstruction of objects in motion. The CWT is also highly robust against sensor noise. Moreover, it is able to handle temporary occlusions resulting from crossing trajectories. These properties are generally not found in techniques rooted in optical flow and block-based motion estimation.

The study of CWTs originally evolved from considering spatio-temporal affine transformations. These were easily amenable to Lie group structures and admissible wavelet representations. The approach turned out to be efficient

for signal analysis and enabled the introduction of numerous physical parameters as criteria of selectivity. The importance of CWTs in this field was recognized several years ago [1]. Although the analysis of image sequences requires numerous analyzing parameters, only a small subset of them has to be considered simultaneously in each specific application. The most significant components of uniform motion are studied in this paper, i.e. the translation, the rotation, the deformation, and the acceleration. The spatial orientation (the preferential axis of inertia) and the scale are additional parameters of concern; indeed, the scale is intrinsic to any wavelet analysis. The application of motion tracking is addressed in this paper and is illustrated with CWTs tuned to the velocity, i.e. the translational motion. It is assumed that local motion is linear. Hence, the technique applies whenever the approximation is valid locally on a few frames (3 or more). CWTs that are tuned to velocities are called *Galilean wavelets* to refer to the Galilei group that describes classical mechanics.

2. BUILDING FAMILIES OF CWTS

The construction of CWTs relies on signal transformations that model motion and object deformations. They can take into account translation, rotation, scale and shear. One idea developed in this work, consists of expressing all these elementary transformations as unitary operators in the spatio-temporal domain $2D + T$ (2D spatial plus time), and to write useful generalizations for uniform motion (i.e. motion described by time-invariant parameters), namely *translation, rotation, deformation*, and *acceleration*. These unitary operators and their related parameters are eventually combined to form a general law of signal transformation. This transformation is intended to be applied either to the signal or to the mother wavelets. When applied to the signal, it describes the transformations performed by the motion i.e. warpings of the signal. When applied to an admissible mother wavelet, it generates a whole family of continuous spatio-temporal wavelets (band pass filters for signal analysis). The construction of these families is ruled by locally compact groups (Lie groups) and the admissibility of a mother wavelet is enforced by three representation properties: square-integrability, irreducibility and unitarity. The procedure to calculate admissible wavelets is well-known and relates to that of the coherent states originating from theoretical physics. As such, it has already been developed in other research work in the one and multi-dimensional cases [2]. The construction proceeds as follows. Unitary spatio-temporal transformations are first defined on the space $(\mathbf{R}^2 \times \mathbf{R})$. Typically, in this construction, the structure of the parameters leads to composition relationships, inverse and identity that characterize a group. The

study of group representations in spatio-temporal Hilbert spaces comes thereafter. According to the theory of coherent states, the demonstration of unitarity, irreducibility and square-integrability of these representations guarantees the existence of admissible continuous spatio-temporal wavelets. Practically, this means that operating the mother wavelet in the space of the parameter definition covers the whole set of bandpass filters of finite energy, while preserving all the well-known wavelet properties (isometry, inversion, reproducing kernel, and resolution of the identity). In this section, five different constructions of CWT families will be considered as examples. They originate from the groups covering all the motions [3] and represent the action of Lie algebras and groups on manifolds. First, operators on wavelet and signal $\{\Omega : L^2(R^2 \times R)\} \rightarrow L^2(R^2 \times R)\}$ will be defined with their respective set of parameters.

The affine-Galilei group supports the construction of CWTs tuned to velocity and *uniform translational motions*. The set of operators and parameters involved in this CWT are the spatio-temporal translation of parameter $\vec{b}$ and τ to represent the space and time locations, the velocity $\vec{v}$, the dilation a to represent the scale, and the spatial rotation θ to represent the orientation (the preferential anisotropy). The action of these parameters can be written as the following spatio-temporal transformation

$$\vec{x}_2 = \frac{1}{a} R(-\theta)(\vec{x}_1 - \vec{b} - \vec{v}t); \quad t_2 = t_1 - \tau , \qquad (1)$$

where $R(\theta)$ is the rotation matrix in SO(2). Let us write the wavelet transform, $\Psi(\vec{x}, t)$, in the Galilean family. In the spatio-temporal domain, we have

$$\begin{aligned}&\left[\Omega(\vec{b}, \tau, \vec{v}, a, \theta)\Psi\right](\vec{x}, t) \\ &= \frac{1}{a}\Psi\left[\frac{1}{a} R(-\theta) (\vec{x} - \vec{b} - \vec{v}t) , t - \tau\right] ,\end{aligned} \qquad (2)$$

and in the Fourier domain, where $\vec{k}$ and ω stand for spatial and temporal frequencies, we have

$$\begin{aligned}&\left[\widehat{\Omega}(\vec{b}, \tau, \vec{v}, a, \theta)\widehat{\Psi}\right](\vec{k}, \omega) \\ &= a \, e^{-i(\vec{k}.\vec{b} + \omega\tau)} \, \widehat{\Psi}\left[R(-\theta) \, a \, \vec{k} , \omega + \vec{k}\vec{v}\right] .\end{aligned} \qquad (3)$$

The set of parameters considered in this family of CWTs is $(\vec{b}, \tau, \vec{v}, a, \theta)$. These CWTs are called Galilean wavelets.

A slightly different approach to Galilean wavelets, called the *kinematical wavelets*, has been described by Duval-Destin and Murenzi [1]. In this case, the set of parameters is $(\vec{b}, \tau, a, c, \theta)$ where a is the spatio-temporal dilation, and c is the speed parameter. c and θ reach the velocity. The spatio-temporal transformation is given

$$\vec{x}_2 = \frac{1}{c^{1/3}a}R(-\theta)(\vec{x}_1 - \vec{b}); \quad t_2 = \frac{c^{2/3}}{a} (t_1 - \tau) \qquad (4)$$

Let us now consider *uniform rotational motion* as a third spatio-temporal transformation, and generate a family of CWTs. Uniform rotational motion is different from the spatial rotation of the SO(2) group in the sense that it incorporates time and space. The resulting velocity is given in this case by

$$\vec{v}(t) = \vec{v}_0 + \vec{\omega} \wedge \vec{x}(t) , \qquad (5)$$

where $\vec{\omega}$ is the angular velocity, $\vec{v}$ is the translational velocity and $\vec{x}(t)$ the current coordinate location of the moving object. The symbol $\wedge$ stands for the cross vector product. Another way of expressing this signal transformation in the image planes is given as

$$\vec{x}_2 = R(-\theta t)\vec{x}_1; \quad t_2 = t_1 - \tau , \qquad (6)$$

where $\left[R(\theta t) = \begin{pmatrix} \cos\theta t & -\sin\theta t \\ \sin\theta t & \cos\theta t \end{pmatrix}\right]$. The set of parameters considered in this CWT family is $(\vec{b}, \tau, a, \vec{v}_0, \vec{\omega})$ or $(\vec{b}, \tau, a, \vec{v}_0, \theta)$.

Uniform temporal dilation (i.e. expansion or contraction) is defined by substituting in Equation (6) $R(\theta t)$ by $\left[D(\alpha t) = \begin{pmatrix} e^{-\alpha t} & 0 \\ 0 & e^{-\alpha t} \end{pmatrix}\right]$. This transformation is important since any object in motion approaching the camera undergoes rather exponential expansions in the image field. The set of parameters of interest for the CWT construction are then $(\vec{b}, \tau, a, \vec{v}_0, \alpha)$.

A fifth set of analyzing parameters would consider *uniform acceleration* $\vec{\gamma}$, given by the second order coefficient when expanding the trajectory curve $\vec{x} = \vec{f}(t)$ in series

$$\vec{x}(t) = \vec{b} + \vec{v}_0 t + \frac{1}{2} \vec{\gamma}_0 t^2 + \sum_{n=1}^{\infty} \frac{1}{(n + 2)!} \vec{\gamma}_n \, t^{n+2} \quad (7)$$

where $\vec{v}_0 = \frac{d\vec{f}(t)}{dt}|_{t=0}$ is the velocity, and $\vec{\gamma}_0 = \frac{d^2\vec{f}(t)}{dt^2}|_{t=0}$ is the acceleration. The $\vec{\gamma}_n$ stands for n^{th}-order acceleration and is not considered in this study. Thus, the parameters of interest in this CWT family will be $(\vec{b}, \tau, a, \vec{v}_0, \vec{\gamma}_0)$.

3. DEFINITION OF THE CWT

This section presents the definition of one CWT family, the *Galilean wavelets*. The signal $s(\vec{x}, t)$ subject to analysis is defined in the Hilbert space $L^2(R^2 \times R, d^2\vec{x}dt)$. The CWT $W[s; \vec{b}, \tau, \vec{v}, a, \theta]$ is defined as an inner product

$$\begin{aligned}W[s; \vec{b}, \tau, \vec{v}, a, \theta] &= c_\Psi^{-1/2} < \Psi_{\vec{b}, \tau, \vec{v}, a, \theta} \mid s > \\ &= c_\Psi^{-1/2} \int_{R^2 \times R} d^2\vec{x} \, dt \, \bar{\Psi}_{\vec{b}, \tau, \vec{v}, a, \theta} \, s(\vec{x}, t)\end{aligned}$$

where the overbar $^-$ stands for the complex conjugate. The wavelet, Ψ, is a *mother wavelet*. It must satisfy the condition of admissibility (i.e. of square-integrability) meaning that there exits a constant c_Ψ such that

$$c_\Psi = (2\pi)^3 \int_{R^2 \times R} d^2\vec{k} \, d\omega \, \frac{|\widehat{\Psi}(\vec{k}, \omega)|^2}{|\vec{k}|^2} < \infty .$$

A numerically efficient way of performing the CWT consists of working in the spectral domain by means of the (2D+T) FFT. The other CWT families have a similar definition.

4. EULER-LAGRANGE EQUATION

Let us consider Lagrange's principle of the least action that can be equivalently derived in classical mechanics and in optimal control from the calculus of variations. The system is characterized by the action S and a non-negative definite function, called the Lagrange function, $L[\vec{x}(t), \dot{\vec{x}}(t); t]$, where $\vec{x}(t)$ is the trajectory and $\dot{\vec{x}}(t) = \frac{d\vec{x}(t)}{dt}$ is the corresponding velocity function. The calculus of variations allows us to derive the motion equation and the trajectory

that optimize the action. Usually, motion between times t_1 and t_2 in a conservative mechanical system coincide with the extremal of the functional

$$S = \int_{t_1}^{t_2} L[\vec{x}(t), \dot{\vec{x}}(t); t] dt , \qquad (8)$$

where L is the difference between the kinetic and the potential energy. Optimal control exploits the same modeling, where S is a cost function to be optimized under some constraints to be specified. The trajectory is then uniquely defined when the initial conditions are known in terms of object location and velocity (detection issue). At the extremum, denoted by $*$, the calculus derives the well-known Euler-Lagrange equation

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\vec{x}}} - \frac{\partial L}{\partial \vec{x}} = 0 . \qquad (9)$$

In this paper, the Lagrange function L to be considered is the square of the modulus of the Galilean CWT, i.e. the energy density $| < \Psi_{\vec{b},\tau,\vec{v},a,\theta} \mid s > |^2$, $\vec{b} = \vec{x}$, $\tau = t$ and $\vec{v} = \dot{\vec{x}}(t)$. The Cauchy-Schwarz inequality states that

$$| \int_{R^2 \times R} d^2\vec{k} \, d\omega \, \Psi \left[\vec{k}, \omega \right] \tilde{s} \left(\vec{k}, \omega \right) |^2$$

$$\leq \int_{R^2 \times R} d^2\vec{k} \, d\omega \, | \Psi \left[\vec{k}, \omega \right] |^2 \int_{R^2 \times R} d^2\vec{k} \, d\omega \, | \tilde{s} \left(\vec{k}, \omega \right) |^2 , \qquad (10)$$

where $\tilde{s}(\vec{x}, t)$ is a band-limited version of $s(\vec{x}, t)$ with one or several moments equal to zero. Then, equality proceeds if $\Psi(\vec{k}, \omega) = c \, \tilde{s}(\vec{k}, \omega)$. This inequality provides some starting conditions for the wavelet transform to perform matched filtering or correlation. The analyzing wavelet has to be matched to the object with respect to its spectrum and its motion. In our case, the unique optimum to be tuned must correspond to the trajectory. This enables a stable and unambiguous tracking procedure. This important property must then be analytically demonstrated for each family of wavelets when applied to the particular motion under investigation. This equation and all its related theory remain valid in our case and interconnect our analysis problem not only to the theory developed for mechanical systems but also to optimum control. The equations and the algorithms that have been developed to recursively construct the optimum control, apply readily to this problem. Let us mention the Kalman filter and Bellman's algorithm (Viterbi algorithm).

5. DETECTION AND TRACKING

The detection of moving objects relies on extracting local maxima in the velocity representation, $E = f(\vec{v}, a)$,

$$E(\vec{v}, a) = \int_{\tau=0}^{\tau=T} \int_{\vec{b}=\vec{b}_{min}}^{\vec{b}=\vec{b}_{max}} | < \Psi_{\vec{b},\tau,\vec{v},a} | s > |^2 \, d\tau d^2\vec{b} \quad (11)$$

i.e. from the energy density computed by integrating the energy of the CWT over the space and the length of the scene. This technique effectively characterizes all the moving objects and the velocities.

The tracking strategy is based on combining *Kalman filters* and *CWTs*. The state of the Kalman filter is composed of all the wavelet parameters. Usually, Kalman filters are characterized by two equations, a *state equation* and an *observation equation*. The state equation is an adaptive predictor that updates the state $U(n)$ of the filter

$$\hat{U}(n) = \Phi(n, n-1)U(n-1) + W(n) , \qquad (12)$$

where $\hat{U}(n)$ is the state prediction at step n and $W(n)$ the prediction error. Φ is the transition matrix or the feedback matrix of the Kalman filter. If the state is well-chosen (i.e. the CWT matches the signal), the predictor behaves as a Markov process, and the prediction error is a zero-mean Gaussian process. In the case of an analysis with Galilean wavelets, the state parameters are composed of the set $(\vec{b}, \tau, \vec{v}, a, \theta)$ and the prediction step n is the image interval. For other CWTs (like the accelerated family), the prediction step can involve several images, typically tens of them. The CWT is then used at each step n as a motion analyzer to determine the exact state values of the Kalman filter $U(n)$. A *gradient algorithm* works in the neighborhood of the predicted state $\hat{U}(n)$ to locate the exact state $U(n)$ composed of the parameters that maximize the following energy density

$$MAX \quad E(\vec{b}, \tau, \vec{v}, a, \theta) = | < \Psi_{\vec{b},\tau,\vec{v},a,\theta} | s > |^2 . \qquad (13)$$

The observation equation also exploits the CWT as a motion-based extraction tool tuned to the current exact state parameters. The CWT captures and isolates the selected objects from the scene s to provide a display I,

$$I(n; \vec{b}, \tau) = < \Psi_{\vec{b},\tau,\vec{v}=\vec{v}_{opt},a=a_{opt},\theta_{opt}} | s > + V(n; \vec{b}, \tau) . \qquad (14)$$

I is the segmented image of the selected object, displayed alone at its correct location; s is the original signal under analysis, and V is the noise produced by the optical sensors.

6. MORLET WAVELET AND APPLICATIONS

The applications presented in this paper for detection and tracking has been performed with the Galilean CWT. An anisotropic *Morlet wavelet* is admissible as a mother wavelet in the Galilean family; it defines a non-separable filter

$$\Phi(\vec{x}, t) = e^{i\vec{k}_0 \vec{X}} e^{-\frac{1}{2} < \vec{X} \mid C\vec{X} >} - e^{-\frac{1}{2} < \vec{k}_0 \mid D\vec{k}_0 >} e^{-\frac{1}{2} < \vec{X} | C\vec{X} >}$$

where $\vec{X} = (\vec{x}, t)^T \in R^n \times R$, C is a positive definite matrix and, $D = C^{-1}$. For $2D + T$ signals, $\left[C = \begin{pmatrix} 1/\epsilon_x & 0 & 0 \\ 0 & 1/\epsilon_y & 0 \\ 0 & 0 & 1/\epsilon_t \end{pmatrix} \right]$ where the ϵ factors introduce anisotropy in the wavelet shape. Figures 1 and 2 show the energy density of the Morlet wavelet in the Fourier domain at velocity $\vec{v} = (1, 0)$. A high selectivity or anisotropy $\epsilon_t = 1000$ has been applied to flatten the wavelet along the velocity plane. Figure 4 presents the issue of the motion detection applied to the synthetic scene displayed in Figure 3. Figures 5 and 6 present the tracking of one accelerated object captured out of five others.

7. CONCLUSIONS

Several families of spatio-temporal CWTs have been proposed in this paper as tools to analyze spatio-temporal signals with respect to mechanical criteria. Among them, the

Galilean wavelet transform is tuned to velocities and uniform translation motion. We have shown how that CWT family can handle detection and tracking applications. We believe, at this point, that the approaches based on CWTs have promise in the area of motion tracking. Tracking has also been shown possible even under severe noise conditions, and even when occlusions occur.

REFERENCES

[1.] M. Duval-Destin and R. Murenzi "Spatio-Temporal Wavelet: Application to the Analysis of Moving Patterns", *in Progress in Wavelets Analysis and Applications (Proc. Toulouse 1992) Y. Meyer and S. Roques, Editors, Ed. Frontières, Gif-sur-Yvette, pp. 399-408, 1993.*

[2.] S.T. Ali, J.-P. Antoine, J.-P. Gazeau "Square integrability of group representations on homogeneous spaces. I. Reproducing triples and frames. II. Coherent and quasi-coherent states. The case of the Poincaré group", *Ann. Inst. H. Poincaré, Vol. 55, pp. 829-855 & 857-890, 1991.*

[3.] D. Martin "Manifold Theory, An introduction for Mathematical Physicists", *Ellis Horwood, England, 1991.*

[4.] L. G. Weiss "Wavelets and Wideband Correlation Processing", *IEEE Signal Processing Magazine, pp. 13-32, January 1994.*

[5.] J.-P. Leduc and C. Labit "Spatio-Temporal Wavelet Transforms for Image Sequence Analysis", VIII European Signal Processing Conference, EUSIPCO-96, Trieste, Italy, 10-13 September 1996, 4 pp.

[6.] J.-P. Leduc "Discrete and Continuous Spatio-Temporal Wavelet transforms", admitted for publication in IEEE Transactions on Signal Processing.
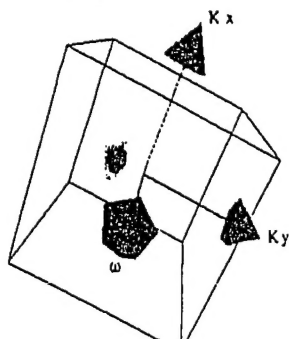
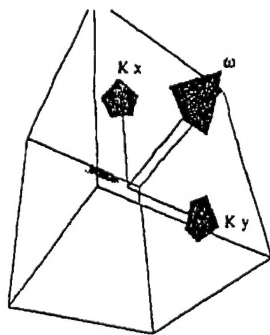Figure 1. Galilean wavelet in velocity plane (1,0).

Figure 2. Galilean wavelet in velocity plane (1,0).

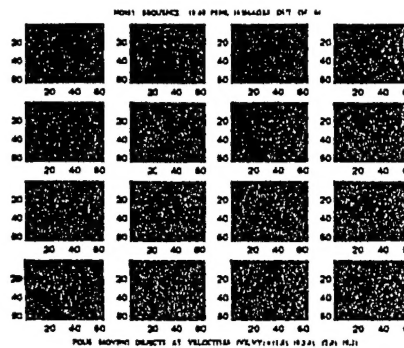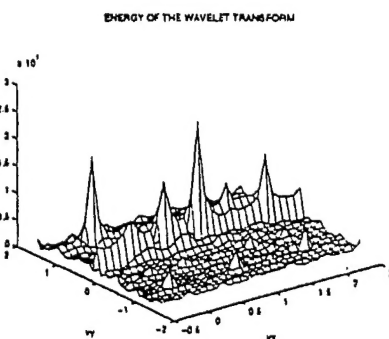

Figure 3. synthetic noisy image sequence.

Figure 4. Velocity detection in the noisy sequence: $\vec{v} = (v_x, v_y) = (0, .5), (0, 1) (0.2), (1, 0)$.
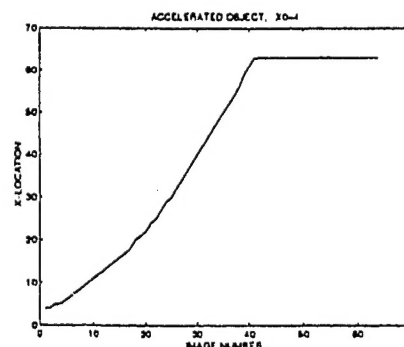

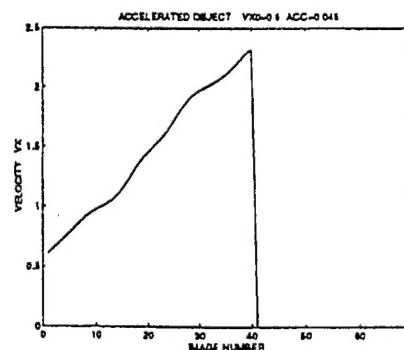Figure 5. Selective trajectory construction (remark: the upper bound image is located at $x = 64$).

Figure 6. Selective velocity tracking.

Appendix B

Spatio-temporal continuous wavelets applied to missile warhead detection and tracking

F. Mujica*, J.-P. Leduc*, and M. J. T. Smith

Center for Signal and Image Processing
School of Electrical Engineering, Georgia Institute of Technology

R. Murenzi

Center for Theoretical Studies of Physical Systems*
Department of Physics, Clark Atlanta University

ABSTRACT

This paper addresses the problem of tracking a ballistic missile warhead. In this scenario, the ballistic missile is assumed to be fragmented into many pieces. The goal of the algorithm presented here is to track the warhead that is among the fragments. It is assumed that images are acquired from an optical sensor located in the interceptor nose cone. This imagery is used by the algorithm to steer the course of interception.

The algorithm proposed in this paper is based on continuous spatio-temporal wavelet transforms (CWTs). Two different energy densities of the CWT are used to perform velocity detection and filtering. Additional post-processing is applied to discriminate among objects traveling at similar velocities. Particular attention is given to achieving robust performance on noisy sensor data and under conditions of temporary occlusions.

First we introduce the spatio-temporal CWT and stress the relationships with classical orientation filters. Then we describe the CWT-based algorithm for target tracking, and present results on synthetically generated sequences.

Keywords: Motion estimation; Target detection and tracking; Continuous wavelet transform

1   INTRODUCTION

One of the more difficult problems related to target tracking is the interception of ballistic missiles. In this scenario, an optical image sensor is located in the nose cone of an interceptor. The images acquired are processed on-board and the course of interception is determined. Several factors complicate the problem. First, the images received are noisy, often making it difficult to isolate the object of interest. Second, ballistic missiles often fragment, resulting in a cloud of many objects traveling on similar trajectories. Among this debris is the target of interest, which is the warhead. Third, it is possible that the trajectories of missile fragments may cross during

the course of their travel. Thus for a brief period of time, the warhead may be occluded by debris. Finally, both the missile and the interceptor are traveling at high speeds. Hence, very little processing delay can be tolerated, since decisions regarding course changes must be made early for a successful interception to occur.

The specific problem we address in this paper is developing an algorithm capable of tracking a target in this noisy environment. It is assumed that the true target location has been passed to the interceptor by a ground station. The algorithm we propose to estimate object trajectories is based on spatio-temporal continuous wavelet transforms (CWTs), and attempts to exploit the multidimensional nature of the representation. The CWT has the attractive property that it is parameterized according to physically meaningful parameters, namely, location $\bar{x}$, time t, scale a, velocity magnitude c, and orientation θ. Multiple feature spaces can be derived by integrating over a subset of the parameters. The structure of the CWT provides a natural way to approach the tracking problem, and has resulted in our developing a three-stage algorithm.

One feature space extracts the velocities embedded in the image sequence. This is called the *velocity detection* stage. In the next stage, a second representation is used to retain those objects moving at a prespecified velocity and to reject background noise and other objects moving at different velocities. Finally, post-processing is performed in order to resolve multiple objects moving at the same velocity.

This problem is largely one of motion estimation and analysis. The most popular techniques of this type are based on *block matching* and *optical flow* schemes. Block matching techniques model motion in terms of a coarse block-translation representation, while *optical flow* methods rely on gradient approximations. These techniques, in general, are highly sensitive to noise [2, 4, 3], and cannot handle occlusions in a natural way.

More appropriate for this application is a filter/transform-based approach that exploits the directional orientation characteristic of the motion in both wave-number[1]/frequency and the spatio-temporal variables. Clever selection and use of directional filters can effectively isolate specific object motion. This approach is more robust to noise and can naturally handle temporary occlusions resulting from crossing trajectories.

Methods using Gabor filters and projection filters have already been considered as techniques to capture the directional characteristic of the motion. However, they generally do not provide sufficient velocity resolution for this application [2]. The spatio-temporal CWT, however, can be effective in this regard. We have obtained excellent velocity estimates from the CWT and achieved rejection of background noise as part of the process. Additionally, the spatio-temporal domain of the CWT allows the interpolation of "missing" information when occlusions occur.

In the following sections, we introduce the spatio-temporal CWT and discuss the relationships with classical orientation filters. A description of the CWT-based algorithm for target tracking is presented next, and supported by experimental results on synthetic data.

2   MOTION ESTIMATION

Motion estimation is a difficult task, especially when noisy imagery and complex motion are considered. Classical algorithms for motion estimation are intrinsically based on approximations of a local gradient, which is typically calculated from a pair of image frames. As such, these techniques are highly sensitive to noise [2, 4, 3]. Another limitation of these techniques is their inability to handle occlusions.

To overcome this problem, a filter/transform-based approach is usually taken. Here the image sequence is considered as the input signal, to take full advantage of time correlations. Among these algorithms are Gabor and projection filtering techniques. Unfortunately, these filters have relatively poor velocity resolution [2, 4], thus,

---

[1]Wave-number in this context refers to the spatial frequencies.

they are not appropriate for the application at hand.

The idea underlying our approach is to process at once as many frames as permitted by memory and processing time restrictions. By expanding the temporal region of support of the filters, we improve robustness against noise and enable complex motion to be handled easily. In addition, "missing" information produced by crossing trajectories can be interpolated, resulting in improved behavior against occlusions. This approach has the drawback of increasing the temporal delay of the overall system, which is of primary importance in the scenario of interest. This implies a tradeoff between accuracy of the estimation and temporal delay. In this paper, we focus on the motion-estimation accuracy.

2.1 DEFINITIONS

The Hilbert space of interest, $\mathcal{H}$, corresponds to the space of square integrable signals over time and space, that is, $L^2(\mathbb{R}^2 \times \mathbb{R}, d^2\vec{x}\,dt)$. The corresponding norm is:

$$||s||^2 = \int_{\mathbb{R}^2} \int_{\mathbb{R}} |s(\vec{x},t)|^2 d^2\vec{x}\,dt \ . \tag{1}$$

To distinguish the dimensions, we use the term (2+1)D to signify two spatial dimensions plus time. The Fourier transform of a (2+1)D signal, $s(\vec{x}, t)$, is denoted by $\hat{s}(\vec{k}, \omega)$, where $\vec{k}$ and ω are the wave-number and temporal frequency respectively.

2.2 MOTION TRANSFORMATIONS ON THE FOURIER SPACE

To understand how motion affects the wave-number/frequency domain consider a steady signal

$$s(\vec{x}, t) = s(\vec{x}) \ . \tag{2}$$

The Fourier transform can be expressed as

$$\hat{s}(\vec{k}, \omega) = \hat{s}(\vec{k})\delta(\omega) \ . \tag{3}$$

The energy is concentrated in the plane defined by $\omega = 0$. Now, consider a moving version (with velocity $\vec{v}$) of the same signal,

$$s(\vec{x} - \vec{v}t, t) \ , \tag{4}$$

and its Fourier transform,

$$\hat{s}(\vec{k}, \omega + \vec{k} \cdot \vec{v}) \ . \tag{5}$$

The energy of the signal is now concentrated in a plane defined by $\omega = -\vec{k} \cdot \vec{v}$ (i.e. a plane perpendicular to the vector $\vec{v}$). Such a plane is called the *velocity plane* associated with the velocity vector $\vec{v}$. The inclination of the resulting velocity plane with respect to the horizontal plane (wave-number plane, $k_x - k_y$) depends only on the magnitude of the velocity, $|\vec{v}|$, while the inclination with respect to the $k_x - \omega$ plane (or $k_y - \omega$ plane) depends on the orientation of the velocity vector, $\alpha = \tan^{-1}(v_y/v_x)$. Figure 1 shows the effect of constant motion on a synthetic signal (Figure 1a) in the wave-number/frequency domain. Three velocity magnitudes are considered, $|\vec{v}| = 0, 1$ and 2 in Figures 1b, c and d respectively.

It is clear that the problem of velocity detection and filtering is a task for orientation selective filtering. The spatio-temporal CWT can be seen as a tool for designing orientation filters in an elegant way. As it will be shown in the subsequent sections, the parameters of the wavelet representation are directly related to motion transformation parameters. These parameters, velocity magnitude and orientation, along with spatial and temporal translations, and scale, define the basic parameters of the spatio-temporal CWT.
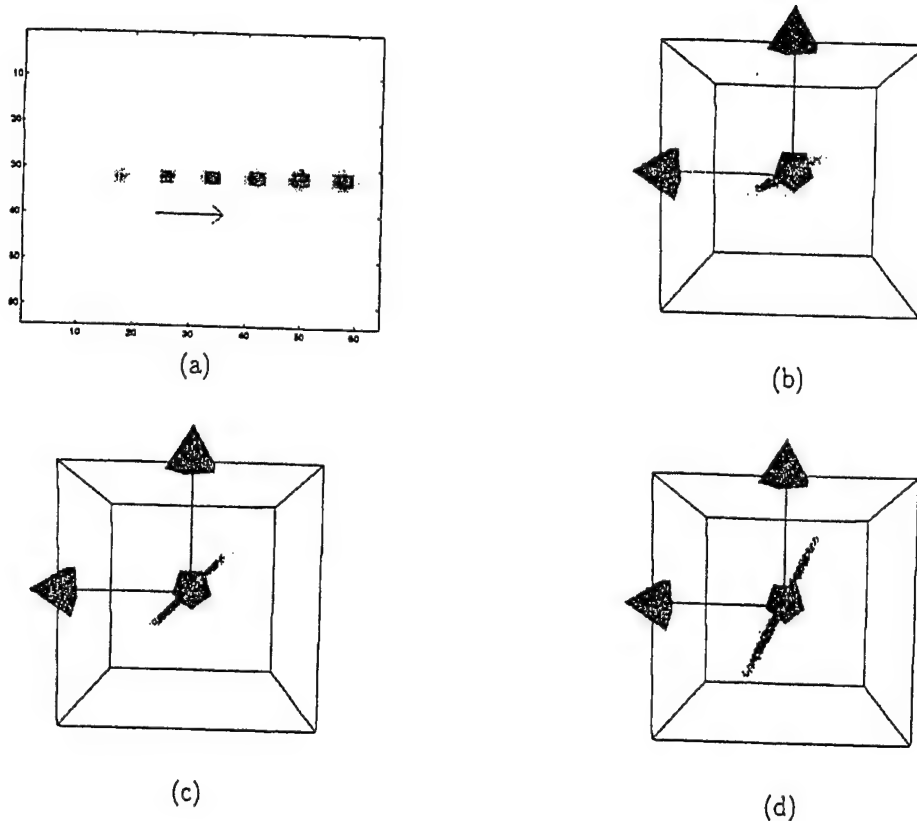
Figure 1: Horizontally moving signal. (a) Space-time domain representation (time evolution is represented by intensity variation) ; (b, c and d) DFT domain representation for $|v| = 0.5, 1.0$ and 2.0 respectively.

3   SPATIO-TEMPORAL CWT

We employ the spatio-temporal CWT in a way that is reminiscent of the operation of the human visual system, in the sense that high spatial resolution is emphasized for slowly moving objects, and high temporal resolution is emphasized for rapidly moving objects [5]. As in the 1D case, the (2+1)D CWT performs a transformation with respect to a set of basis functions that are derived from a particular signal, called the mother wavelet, by means of function transformations. In the next section, the transformations applied to the *mother wavelet*, ψ, both in the original and in the transformed domain, are briefly described.

3.1   TRANSFORMATIONS APPLIED TO THE WAVELET

Let us assume the region of support of the mother wavelet is concentrated in a velocity plane defined by $\omega = -\vec{k} \cdot \vec{v}_o$. The condition that the signal ψ must satisfy to be considered a mother wavelet, and the definition of the one used in this work will be described in a subsequent section.

The function transformation used to derive the set of bases for the (2 + 1)D CWT are as follows [5, 7].

- Spatial and temporal translation: The wavelet is shifted to a given point in space-time. This transformation

is denoted by $T^{(\vec{b},\tau)}$ and is defined by

$$\left[T^{(\vec{b},\tau)}\psi\right](\vec{x},t) = \psi(\vec{x}-\vec{b},t-\tau) \tag{6}$$

$$\left[T^{(\vec{b},\tau)}\hat{\psi}\right](\vec{k},\omega) = e^{-j(\vec{k}\cdot\vec{b}+\omega\tau)}\hat{\psi}(\vec{k},\omega) . \tag{7}$$

- Rotation: This transformation, denoted as R^θ, rotates the wavelet on the spatial coordinates around the frequency axis. In this way the filters can be tuned to a particular orientation associated with that velocity. This transformation is defined by

$$\left[R^\theta\psi\right](\vec{x},t) = \psi(r^{-\theta}\vec{x},t) , \tag{8}$$

$$\left[R^\theta\hat{\psi}\right](\vec{k},\omega) = \hat{\psi}(r^{-\theta}\vec{k},\omega) , \tag{9}$$

$$r^{-\theta} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} . \tag{10}$$

- Scaling: This transformation, denoted by D^a, is standard in the wavelet framework and is defined by

$$[D^a\psi](\vec{x},t) = a^{-3/2}\psi(\frac{\vec{x}}{a},\frac{t}{a}) , \tag{11}$$

$$\left[D^a\hat{\psi}\right](\vec{k},\omega) = a^{3/2}\hat{\psi}(a\vec{k},a\omega) . \tag{12}$$

- Speed tuning: This transformation can be seen as two independent scaling operations performed on the spatial and temporal variates. This allows for the localization of the wavelet around the correct inclination of the velocity plane. It is denoted by Λ^c and defined by

$$[\Lambda^c\psi](\vec{x},t) = \psi(c^{-1/3}\vec{x},c^{2/3}t) , \tag{13}$$

$$\left[\Lambda^c\hat{\psi}\right](\vec{k},\omega) = \hat{\psi}(c^{1/3}\vec{k},c^{-2/3}\omega) . \tag{14}$$

Thus, the parameter c is directly associated with the velocity magnitude, $|\vec{v}|$. This parameter was introduced in [5], in order to resolve the tradeoff between high (low) spatial frequency and low (high) temporal frequency for slow (fast) moving objects.

These operators taken all together transform a particular signal with region of support around location $(a,c,\theta,\vec{b},\tau)$ in the parameter space, to a new location $(a',c',\theta',\vec{b}',\tau')$, provided the mother wavelet has appropriate support.

3.2 DEFINITION

The spatio-temporal CWT of a (2+1)D signal is a mapping from finite energy signals defined on $(\mathbb{R}^2 \times \mathbb{R})$ to finite energy signals defined on

$$G = g : (a,c,\theta,\vec{b},\tau) \in (\mathbb{R}^+ \times \mathbb{R}^+ \times [0,2\pi] \times \mathbb{R}^2 \times \mathbb{R}) . \tag{15}$$

Consider the composite transformation defined by the application of all the operators defined in the previous section,

$$[\Omega_g\psi](\vec{x},t) = \left[T^{\vec{b},\tau}R^\theta\Lambda^c D^a\psi\right](\vec{x},t) \tag{16}$$

$$= a^{-3/2}\psi(\frac{1}{c^{1/3}a}r^{-\theta}(\vec{x}-\vec{b}), \frac{c^{2/3}}{a}(t-\tau)) , \tag{17}$$

$$\left[\hat{\Omega}_g\hat{\psi}\right](\vec{k},\omega) = \left[T^{\vec{b},\tau}R^\theta\Lambda^c D^a\psi\right](\vec{k},\omega) \tag{18}$$

$$= a^{3/2}\hat{\psi}(ac^{1/3}r^{-\theta}\vec{k}, \frac{1}{ac^{2/3}}\omega)e^{-j(\vec{k}\cdot\vec{b}+\omega\tau)} . \tag{19}$$

For convenience, let us define $\psi_{(a,c,\theta,\vec{b},\tau)}(\vec{x},t) = [\Omega_g\psi](\vec{x},t)$ and its equivalent Fourier transform to be, $\hat{\psi}_{(a,c,\theta,\vec{b},\tau)}(\vec{k},\omega) = \left[\hat{\Omega}_g\hat{\psi}\right](\vec{k},\omega)$. The spatio-temporal CWT of a $(2+1)$D signal, s, with respect to the mother wavelet, ψ, is defined by [5]

$$S_\psi(a,c,\theta,\vec{b},\tau) = \frac{1}{\sqrt{c_\psi}}\langle\psi_{(a,c,\theta,\vec{b},\tau)}|s\rangle$$

$$= \frac{1}{\sqrt{c_\psi}}\int\int \psi^*_{(a,c,\theta,\vec{b},\tau)}(\vec{x},t)s(\vec{x},t)d^2\vec{x}\, dt . \tag{20}$$

The speed parameter, c, controls the weighting between spatial and temporal planes. Different orientations are considered by applying a rotation operator, with parameter θ, on the spatial coordinates $\vec{x}$. The other parameters, a, $\vec{b}$, and τ, represent scale, spatial translation and temporal translation. The condition for the signal ψ to be considered as a mother wavelet, is called the admissibility condition, and reads [5]

$$c_\psi = (2\pi)^3 \int_{\mathbb{R}^2}\int_{\mathbb{R}} \frac{|\hat{\psi}(\vec{k},\omega)|^2}{|\vec{k}|^2|\omega|}d^2\vec{k}\, d\omega < \infty . \tag{21}$$

In the wave-number/frequency domain, (the $\vec{k}/\omega$ domain), the spatio-temporal CWT may be expressed as

$$S(a,c,\theta,\vec{b},\tau) = \frac{1}{\sqrt{c_\psi}}\langle\hat{\psi}_{(a,c,\theta,\vec{b},\tau)}|\hat{s}\rangle . \tag{22}$$

Here ψ is normalized to $c_\psi = 1$. The spatio-temporal CWT satisfies Parseval's condition of energy conservation,

$$\xi = \int\int\int\int\int |S(a,c,\theta,\vec{b},\tau)|^2\frac{da}{a^4}\frac{dc}{c}\, d\theta\, d^2\vec{b}\, d\tau$$

$$= ||\tilde{s}||^2 = ||s||^2 . \tag{23}$$

Thus, $E \equiv |S(a,c,\theta,\vec{b},\tau)|^2$ is the energy density in the CWT domain.

The energy in the wave-number/frequency domain of a 2D signal moving with constant velocity, $\vec{v}$, is concentrated in a *velocity plane* defined by $\omega = -\vec{k}\cdot\vec{v}$. Proper selection of the CWT parameters can achieve velocity selective filters (i.e. filters with regions of support close to a particular velocity plane). Thus, the CWT represents a suitable tool for velocity detection and filtering. Nevertheless, selection of mother wavelets remains a key issue, since usefulness of the representations depends on that choice. The Mother wavelet used in the simulations is described in the next section.

3.3 MORLET WAVELET

The spatio-temporal extension of the Morlet wavelet is a good candidate for motion estimation applications. Its region of support can be appropriately located around a particular velocity plane, $\vec{v}_o$, typically $\vec{v}_o = [1,0]'$.

Moreover an anisotropy parameter, ϵ, applied to the spatial variates controls the variance of the wavelet with respect to the reference velocity plane. The Morlet wavelet is defined, in the spatio-temporal domain, by

$$\psi(\vec{x}, t) = (e^{j\vec{k}_o \cdot \vec{x}} e^{-\frac{1}{2}|\vec{x}|^2} - e^{-\frac{1}{2}|\vec{x}|^2} e^{-\frac{1}{2}|\vec{k}_o|^2})(e^{j\omega_o t} e^{-\frac{1}{2}t^2} - e^{-\frac{1}{2}t^2} e^{-\frac{1}{2}\omega_o^2}) , \tag{24}$$

and in the wave-number/frequency domain by,

$$\hat{\psi}(\vec{k}, \omega) = (e^{-\frac{1}{2}|\vec{k} - \vec{k}_o|^2} - e^{-\frac{1}{2}(|\vec{k}|^2 + |\vec{k}_o|^2)})(e^{-\frac{1}{2}(\omega - \omega_o)^2} - e^{-\frac{1}{2}(\omega^2 + \omega_o^2)}) . \tag{25}$$

Figure 2 shows the transformed Morlet wavelet, $\hat{\psi}_{a,c,\theta,\vec{\delta},\tau}(\vec{k}, \omega)$, in the wave-number/frequency domain. The three filters shown (two views of each) are tuned to the velocities of the objects from Figure 1, that is, for $c = 0.5, 1.0$ and 2.0. The other parameters are fix as follow, $\theta = 0, \vec{k}_o = [6,0]', \omega_o = 6$ and $\epsilon = .25$. A set of scales, $a = \{1, 1.25, 1.5, 1.75, 2\}$, is chosen to cover uniformly a given velocity plane. As it can be seen from Figure 2,
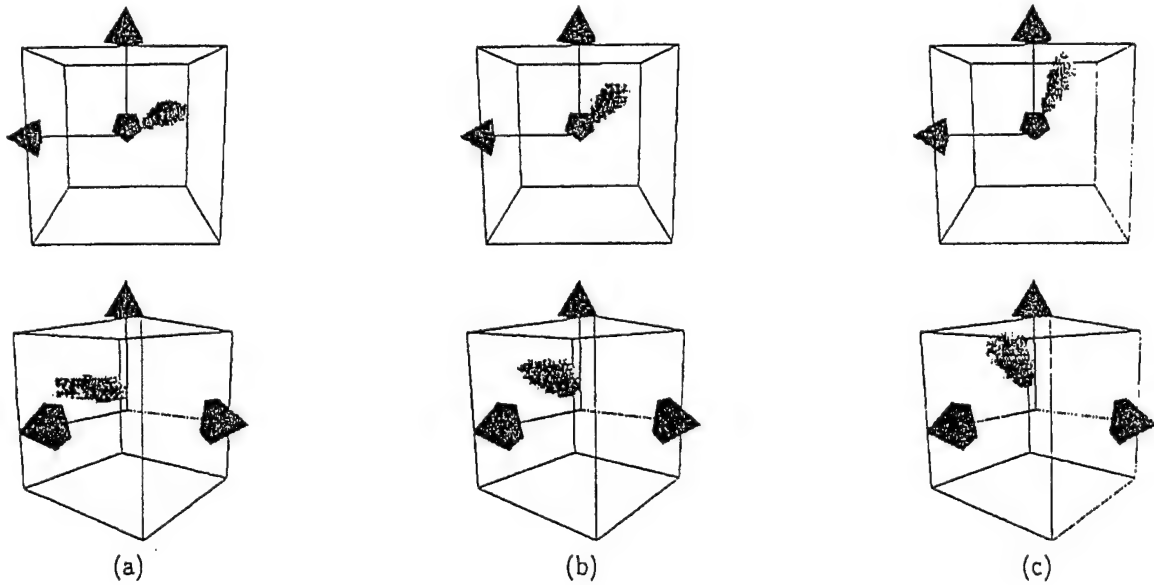

(a)   (b)   (c)

Figure 2: Wave-number/frequency representation of three Morlet wavelet filters with parameters $\theta = 0, \vec{k}_o = [6,0]', \omega_o = 6, \epsilon = 0.25$ and $a = \{1.0, 1.25, 1.5, 1.75, 2.0\}$. (a) $c = 0.5$, (b) $c = 1.0$, (c) $c = 2.0$.

the region of support of the filters is an ellipsoidal cone concentrated around a particular velocity plane. Thus, velocity detection and filtering is possible.

3.4   ENERGY DENSITIES

The multidimensional nature of the CWT calls for a reduction of the variates to consider in the visualization process. This can be done by either fixing a subset of the parameters, or by integrating over a subspace of the parameter space. The second approach has the nice property that it can result in invariant representations with respect to some parameters. Thus, partial integration on a subset of the parameters in the energy density results in different energy representations that can be used to extract relevant features. Among the various possibilities for energy representations, two are of particular interest for object tracking and detection applications.

- Speed-orientation energy density: Here integration is performed over scale, a, spatial translation. $\vec{b}$, and temporal translation, τ. As a result, we obtain the following energy density,

$$E_1(c, \theta) = \int_{\mathbb{R}^+} \int_{\mathbb{R}^2} \int_{\mathbb{R}} \frac{1}{c} \left| < \psi_{a,c,\theta,\vec{b},\tau} | s > \right|^2 \frac{da}{a^4} d^2 \vec{b} \, d\tau \; . \tag{26}$$

- Space-time-velocity energy density: In this representation, the speed tuning parameter and the orientation are fixed ($c = c_i, \theta = \theta_i$) and integration over the scale parameter, a, is performed. The resulting energy density is given by,

$$E_2(\vec{b}, \tau, c_i, \theta_i) = \int_{\mathbb{R}^+} \frac{1}{c} \left| < \psi_{a,c_i,\theta_i,\vec{b},\tau} | s > \right|^2 \frac{da}{a^4} \; . \tag{27}$$

The impact of these two energy densities on the proposed detection and tracking algorithm is discussed in the next section.

4   DETECTION AND TRACKING ALGORITHM

At this stage, it is important to mention that a computer implementation of the CWT requires a discretization of its parameters. Thus, integration in the formulas described in the previous sections should be replaced by appropriate summations. In this section, we redefine the two energy representations E_1 and E_2 of equations (26) and (27) for the discrete case.

The detection and tracking algorithm relies on the multidimensional nature of the spatio-temporal CWT [5]. Two different energy density representations are used to estimate motion parameters. The first two stages of the proposed algorithm are depicted in Figure 3.
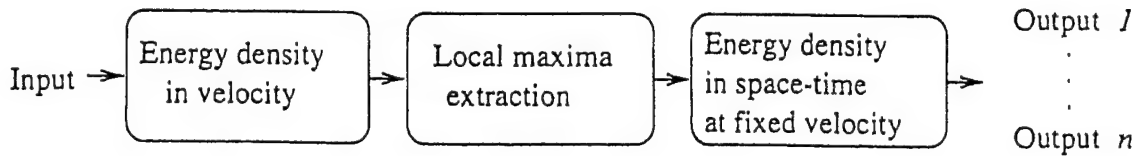

Figure 3: Object Detection and Tracking Algorithm.

The first stage, *velocity detection*, calculates the energy density in the *speed-orientation* representation.

$$E_1(c, \theta) = \sum_{a_j, \vec{b}_{n,m}, \tau_i} \frac{1}{a_j^4 c} \left| < \psi_{a_j, c, \theta, \vec{b}_{n,m}, \tau_i} | s > \right|^2 \; . \tag{28}$$

Here local maxima (c_i, θ_i) correspond to velocities present in the scene. These local maxima are used in the next stage, *velocity filtering*, to calculate the energy density in the *space-time* representation for each pair (c_i, θ_i),

$$E_2(\vec{b}, \tau, c_i, \theta_i) = \sum_{a_j} \frac{1}{a_j^4 c_i} \left| < \psi_{a_j, c_i, \theta_i, \vec{b}, \tau} | s > \right|^2 \; . \tag{29}$$

The result is a series of image sequences, each containing objects moving at a given velocity. Other local features, like shape, scale and orientation, can be used to discern between objects moving at the same velocity. This algorithm provides a means for motion-based detection and tracking in a self contained manner.

5   RESULTS AND CONCLUSIONS

Algorithm development and testing were performed using synthetically generated sequences. In this initial work, we restrict the analysis to linear motion. Figure 4a shows a $64 \times 64 \times 64$ pixel representation of four moving objects. The velocities of each of the four objects are specified in the following table:
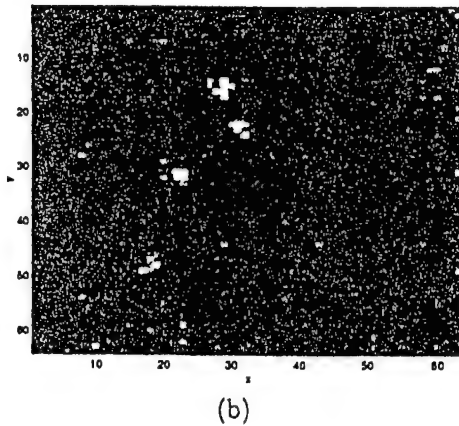
Object	v_x	v_y
1	0.5	0
2	1	0
3	2	0
4	0	1

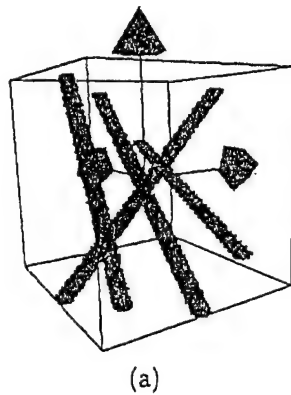(a)   (b)

Figure 4: Four-object test sequence in original domain.

Note that object 4 occludes object 2 around the middle of the sequence. Gaussian noise [2] is added to the sequence to simulate realistic imagery. Frame number 20 of the noisy sequence is shown in Figure 4b.

The result of the *velocity detection* stage (i.e. energy density in the speed-rotation representation) is shown in Figure 5a. The local maxima are represented as plus signs in the contour plot. These local maxima are used in the *velocity filtering* stage. Frame number 20 of the resulting output sequences are shown in Figure 5c.

As a result, each sequence contains just objects moving at a given velocity, and the background noise has been reduced. In the *post-processing* stage, thresholding and local maxima extraction is used to recover the trajectories of all four objects. Figure 5b shows the trajectories obtained. Note that trajectories are correctly interpolated when occlusion occurs. It is worth mentioning that even though no explicit motion model is assumed, trajectories are accurately found.

At this point, we believe the CWT-based method has promise in the area of motion tracking. Velocities were successfully detected and velocity filtering was demonstrated in a natural way, leading to noise reduction in the output. Tracking of the object trajectories was shown to be possible even under severe noise conditions, and even when occlusions were present. Work continues in the CWT development aimed at addressing complex motion.

---

[2]The resulting signal has a SNR of 10 dB with respect to the original signal.
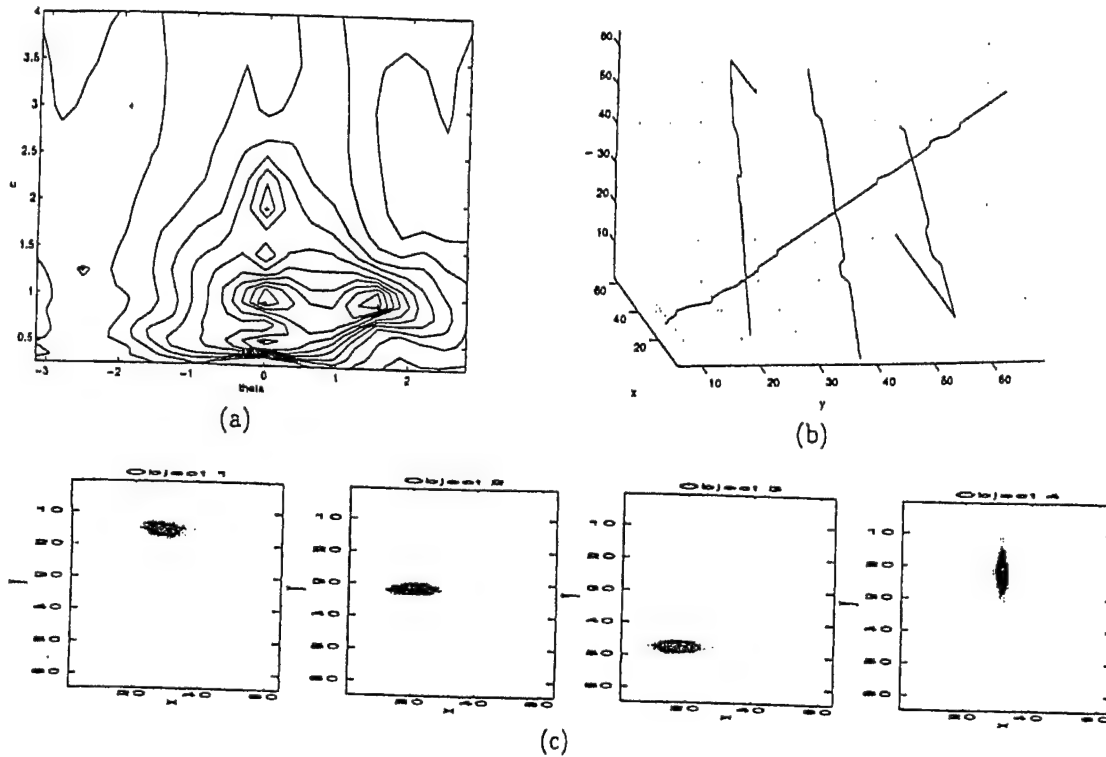
(a)

(b)

(c)

Figure 5: Detection and tracking algorithm results. (a) Velocity detection stage results, E_1; (b) Trajectories of the four objects; (c) Frame number 20 of the filtering stage result, E_2.

ACKNOWLEDGMENTS

APPENDIX

Here the spatio-temporal CWT in the temporal domain, equation (20), is calculated for a moving impulse,

$$s(\vec{x} - \vec{v}t) \ . \tag{30}$$

A simplified version of the Morlet wavelet is used, $|\vec{k}_o| > 5$ and $\omega_o > 5$ [5]. This Morlet wavelet is defined by,

$$\psi(\vec{x}, t) = e^{-\frac{1}{2}(|\vec{x}|^2 + t^2)} e^{-j(\vec{k}_o \cdot \vec{x} + \omega_o t)} \ . \tag{31}$$

After applying all the operators defined by equation (19), we obtain the set of basis functions defined by,

$$\psi_{(a,c,\theta,\vec{b},\tau)}(\vec{x}, t) = a^{-3/2} e^{-\frac{1}{2a^2 c^{2/3}}|\vec{x} - \vec{b}|^2} e^{\frac{-c^{4/3}}{2a^2}(t-\tau)^2} e^{\frac{-j}{ac^{1/3}}\vec{k}_o \cdot r^\theta(\vec{x} - \vec{b})} e^{\frac{-jc^{2/3}\omega_o}{a}(t-\tau)} \ . \tag{32}$$

It can be demonstrated that the CWT of the signal, s, is given by,

$$S_\psi(a,c,\theta,\vec{b},\tau) = \sqrt{2\pi}a^{-3/2}\sqrt{\frac{ac^{2/3}}{c^2+|\vec{v}|^2}}e^{-\frac{1}{2}f_1(a,c,\theta,\vec{b},\tau,\vec{v},\vec{k}_o,\omega_o)}e^{jf_2(a,c,\theta,\vec{b},\tau,\vec{v},\vec{k}_o,\omega_o)} ,\qquad (33)$$

where the functions f_1 and f_2 are defined by,

$$f_1(a,c,\theta,\vec{b},\tau,\vec{v},\vec{k}_o,\omega_o) = \frac{|\vec{b}|^2}{a^2c^{2/3}} + \frac{c^{4/3}\tau^2}{a^2} - \frac{(c^2\tau+\vec{b}\cdot\vec{v})^2}{a^2c^{2/3}(c^2+|\vec{v}|^2)} + \frac{(c\omega_o+\vec{k}_o\cdot r^\theta\vec{v})^2}{c^2+|\vec{v}|^2} ,\qquad (34)$$

$$f_2(a,c,\theta,\vec{b},\tau,\vec{v},\vec{k}_o,\omega_o) = \frac{c^{2/3}\tau\omega_o}{a} + \frac{\vec{k}_o\cdot r^\theta\vec{b}}{ac^{1/3}} - \frac{(c^2\tau+\vec{b}\cdot v)(c\omega_o+\vec{k}_o\cdot r^\theta\vec{v})}{ac^{1/3}(c^2+|\vec{v}|^2)} .\qquad (35)$$

Consider the spatio-temporal CWT of the moving impulse as a function of its velocity, $\vec{v}$, the speed parameter, c, and the orientation parameter, θ, that is,

$$S_{\vec{v}}(a,c,\theta,\vec{b},\tau) .\qquad (36)$$

We expect to obtain a maximum of this expression when,

$$c = |\vec{v}|^2 = \frac{|\vec{b}_l|^2}{\tau_l}\qquad (37)$$

$$\theta = \tan^{-1}(\frac{v_y}{v_x}) = \tan^{-1}(\frac{b_{y_l}}{b_{x_l}}) .\qquad (38)$$

Figure 6a shows the evaluation of equation (36) for $\theta = 0, \vec{b} = [1,0]'$, $\tau = 1$. The Morlet wavelet has parameters $\vec{k}_o = [6,0]'$ and $\omega_o = 6$. The maximum is located around $c = v_x = 1$. Indeed an object with horizontal velocity $v_x = 1$ gets to the point $\vec{b} = [1,0]'$ at $\tau = 1$. In Figure 6b, this experiment is repeated for $\vec{b} = [0,2]'$ and $\tau = 1$. In this case, the peak is found around $c = v_y = 2$.
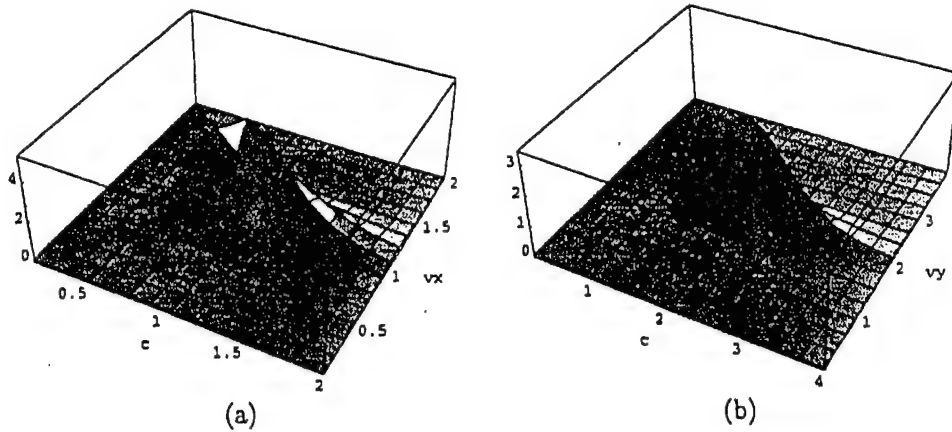

(a)   (b)

Figure 6: Spatio-temporal CWT of a moving impulse as a function of v_x (or v_y) and c for fixed $a, \theta, \vec{b}$, and τ; (a) $S_{\vec{v}=[v_x,0]'}(a = .1, c, \theta = 0, \vec{b} = [1,0]', \tau = 1)$, (b) $S_{\vec{v}=[0,v_y]'}(a = .1, c, \theta = \pi/2, \vec{b} = [0,2]', \tau = 1)$.

6 REFERENCES

[1] Yaakov Bar-Shalom and Thomas E. Fortmann, *Tracking and Data Association*, Academic Press, 1988.

[2] Bernd Jähne, *Digital Image Processing*, Springer-Verlag, 1995.

[3] Thomas J. Burns, Steven K. Rogers, Dennis W. Ruck, and Mark E. Oxley, "Computing Optical Flow Using a Discrete, Spatio-Temporal, Wavelet Multiresolution Analysis", *SPIE Wavelet Applications*, Proc. 2242, pp. 549–560, 1994.

[4] A. Murat Tekalp, *Digital Video Processing*, Prentice Hall, 1995.

[5] M. Duval-Destin and R. Murenzi, "Spatio Temporal Wavelets", in *Progress in Wavelets Analysis and Applications (Proc. Toulouse 1992)*, Y. Meyer and S. Roques, Editors, pp. 399–408, Ed. Frontières, 1993.

[6] K. Nishiguchi, M. Kobayashi, and A. Ichicawa, "Small Target Detection from Image Sequences Using Recursive Max Filter", *SPIE Signal and Data Processing of Small Targets*, Proc. 2561, pp. 153–166, 1995.

[7] J-P Antoine and R. Murenzi, "The Continuous Wavelet Transform, From 1 to 3 Dimensions", in *Subband and Wavelet Transforms: Design and Applications*, Ali N. Akansu and Mark J. Smith, Editors, Kluwer Academic Publishers, pp. 149–187, 1995.

Appendix C

Sensor data enhancement of ballistic missile warheads using an affine motion model

Joseph W. Monaco and Mark J. T. Smith

Center for Signal and Image Processing (CSIP)
Department of Electrical and Computer Engineering
Georgia Institute of Technology

ABSTRACT

Neutralizing the threat of an incoming ballistic missile is a difficult task. Often the missile disintegrates, leaving the warhead surrounded by a number of ballistic fragments. Among these fragments only the warhead must be intercepted. Thus the challenge is for the interceptor to identify and track the warhead so that a successful strike can be achieved. This paper addresses the problem of using noisy image sequence data captured by on-board sensors in the nose cone of the interceptor to detect warheads among fragments. Detection and tracking necessitates the extraction of reliable motion estimates, which is often difficult when the data sequence is noisy.[1] We propose a simple algorithm that exploits both spatial and temporal correlation to suppress noise in the observed image sequence. Our algorithm calculates reliable estimates of the global motion caused by camera movement and uses these motion estimates to reduce significantly the noise in the video sequence.

Keywords: target tracking, noise suppression, affine motion estimation

1 INTRODUCTION

A variety of simple block and optical flow based motion estimation algorithms are used for target tracking and detection problems in scenarios involving imaging sensors in high speed missiles.[2] In general, optical flow methods have an advantage in accurately representing the true motion at the expense of high computational complexity, while simple block-based techniques are faster but do not handle dilation and rotational motion found in high speed missile imagery. In either case, the tracking or detection problem is complicated by the low signal-to-noise ratios that are characteristic of these types of image sequences. As a result, spatial processing in often applied as a preprocessing step to suppress noise.[2]

We propose a better preprocessing step that combines temporal information with spatial processing to create a spatio-temporal noise suppression algorithm. The key to this approach is using an affine motion model to design the spatio-temporal filter. This affine motion model provides more flexibility than traditional translation motion models while at the same time requiring less computation than typical pel-recursive motion estimation techniques. In addition to being useful for noise suppression, the parameters of the affine model can also serve as initial estimates in subsequent tracking algorithms.
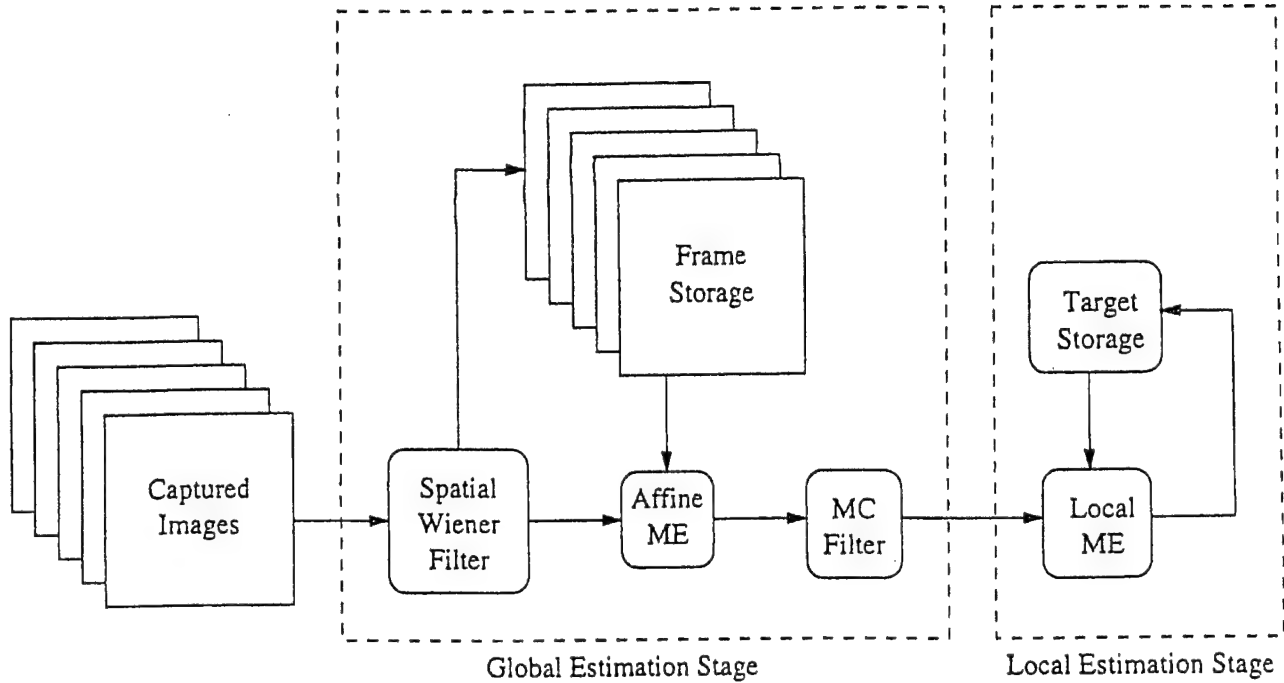
Figure 1: Block diagrams of ballistic missile tracking algorithm

A block diagram of the proposed noise suppression algorithm is shown in Figure 1. This preprocessing step is envisioned to be an important component in the complete tracking algorithm. The observed image sequence is captured by the interceptor as it closes in on the target. The most important task in this noise suppression algorithm is accurate estimation and representation of the global motion created by the movement of the camera. Three issues are of primary importance in designing the motion estimation procedure: flexibility in the motion model, robustness to noise, and computational complexity. Each of these issues is addressed in the following sections.

2   GENERALIZED MOTION MODEL

Since the missile approaches the debris field at a high rate of speed, the model used to represent the global motion field must allow rotation and scale changes. With this requirement in mind, consider a generalized motion model that allows arbitrarily complex motion. The relationship between two frames at time t and $t + \delta t$ is described by:

$$X(n_1, n_2, t) = X(n_1 + \overline{\alpha}^T \overline{\theta}(n_1, n_2), n_2 + \overline{\beta}^T \overline{\phi}(n_1, n_2), t + \delta t), \tag{1}$$

where $X(n_1, n_2, t)$ represents the video signal luminance and $\overline{\theta}(n_1, n_2) = [\theta_1(n_1, n_2)...\theta_p(n_1, n_2)]^T$ and $\overline{\phi}(n_1, n_2) = [\phi_1(n_1, n_2)...\phi_q(n_1, n_2)]^T$ are basis functions used to represent the motion field. The goal of the motion estimation algorithm is to find optimal values for the coefficients $\overline{\alpha}$ and $\overline{\beta}$. A detailed description of the parameter optimization algorithm used for computing $\overline{\alpha}$ and $\overline{\beta}$ has been previously described so that only an overview is presented below.[3] The iterative algorithm is summarized by the following flowchart:

1. Initialization: Set iteration counter $n = 0$ and initialize parameter vectors $\overline{\alpha}_0$ and $\overline{\beta}_0$. Evaluate error function with initial parameter vector and store the result.
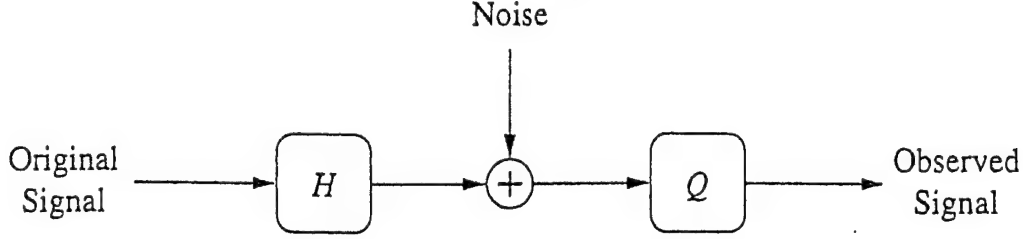
Figure 2: Model for observed signal

2. Taylor Series Expansion: Calculate derivatives at $I(x + \overline{\alpha}_n^T \overline{\theta}(x,y), y + \overline{\beta}_n^T \overline{\phi}(x,y)) \forall (x,y) \in S$

3. Update: Minimize the approximated error function and update parameter vectors to $\overline{\alpha}_{n+1}$ and $\overline{\beta}_{n+1}$.

4. Evaluation: Evaluate the *true* error function. Store parameter error if error function decreases.

5. Termination test: Evaluate stopping criterion. Increment n and return to step 2 if convergence not yet reached.

The approximated error function minimized in Step 3 of the algorithm is given by:

$$\sum_{[n_1,n_2]} \sum W(n_1,n_2) \left(I(n_1,n_2,t) - I(n_1,n_2,t+\delta t) - \frac{\partial I(n_1,n_2)}{\partial n_1} \overline{\alpha}^T \overline{\theta}(n_1,n_2) - \frac{\partial I(n_1,n_2)}{\partial n_2} \overline{\beta}^T \overline{\phi}(n_1,n_2) \right). \quad (2)$$

Likewise, the basis functions are selected as follows to allow an affine transformation:

$$\overline{\theta}(n_1,n_2) = \overline{\phi}(n_1,n_2) = \begin{bmatrix} 1 \\ n_1 \\ n_2 \end{bmatrix}. \quad (3)$$

An affine function was chosen in this case because it meets the requirements of allowing rotation and scale. If a more complex global motion field is expected, the same parameter estimation algorithm can be used.

The first point worth noting is the estimation process relies on the image gradient. This fact increases sensitivity to noise, which is clearly undesirable. We overcome this limitation by applying a spatial filter described in Section 3 to all images before attempting to estimate global motion parameters. Additional robustness is also provided by using the entire field of debris when estimating parameters rather than using a single target. Using the entire debris field implies that many pixels at different location are integrated into the estimation process. Another point worth noting about the estimate algorithm is that it assumes linear variation of the intensity function. Again, this assumption becomes less accurate in noisy environments, but the spatial filter applied to each image tends to enforce the linearity constraint.

3 ROBUSTNESS TO NOISE

The parameter estimation procedure presented in Section 2 degrades in the presence of noise. Fortunately, a spatial filter can be used to reduce the noise while still preserving enough of the original signal energy to calculate global motion parameters. The observed video sequence is modeled as shown Figure 2. Noise is added to the signal after the original image sequence is blurred with the known point spread function $H(w_1,w_2)$. The noise component is assumed to consist of uncorrelated Gaussian noise with a known variance σ^2. A more accurate

Symbol	Operation
T_g	Gradient
T_s	Sample image on non-integer lattice
T_a	Arithmetic operation (add/multiply/compare)
T_i	Matrix inverse

Table 1: Abbreviations used to count floating point operations

noise model could also account for degradations in signal quality caused by clouds or other variables. The final component of the observation model is a quantizer that accounts for the fact that the sensor has finite dynamic range and fixed resolution.

Given these assumptions, an appropriate choice for the initial spatial filter is a Wiener filter. Since the statistics of the original sequence are unknown, the filter used in our simulations is given by:

$$G(w_1, w_2) = \frac{H^*(w_1, w_2)}{|H(w_1, w_2)|^2 + \sigma^2}. \tag{4}$$

This low-pass filter simultaneously accomplishes the goals of reducing noise and imposing linearity on the luminance function. Object boundaries will be blurred with this spatial filter, but the global motion parameters can still be estimated.

4 COMPUTATIONAL COMPLEXITY

The number of floating point operations per second required to calculate the global motion parameters using the generalized motion model is estimated in this section. Computation complexity is of critical importance in the ballistic missile setting because a cost-effective solution that can handle real time video is essential. Since the estimation procedure is iterative, the first obvious question concerns convergence. Although convergence cannot be analytically proven with this approach, our experience in a variety of settings has led to the conclusion that at most five updates are needed for each frame. Another important issue in iterative algorithms is finding an appropriate starting point. We address this problem by initializing all parameters to zero in the first frame. In subsequent frames, the parameters obtained by processing the previous frame serve as initial estimates in the current frame. This approach eliminates the need for any computation to be used in finding starting points in each frame. Table 1 defines the abbreviations used below in computing the total number of operations. A conservative estimate for the number of calculations required in each stage of the algorithm to process an $N \times M$ image is given as follows:

1. Weight function calculation: Many pixels in the image do not provide any motion information. Furthermore, since the motion estimation algorithm is based on the optical flow equation, only areas around edges prove useful for estimating motion.[4] Therefore, we generate the weight function by setting the weight to zero for any pixel whose value is below T_1 or any pixel whose gradient is below T_2. Use of this weight function drastically reduces the computation required in the complete algorithm. The complexity required to compute this mask is given by $NM(T_g + T_a)$.

2. Error function evaluation: In order to evaluate the error function, the previous image must be resampled at the positions given in Equation (1). A fast scan-line algorithm could be used to speed up the computation of image coordinates, but the approximation given here assumes the coordinates are computed directly for each pixel in the image.[5] Therefore, the complexity introduced by evaluating the error function is given by $p(11T_a + T_s)$, where p represents the number of non-zero values in the weight matrix.

3. Parameter update: Updating the motion parameters for each iteration requires a 6×6 matrix inverse. In addition, building the matrix that eventually gets inverted requires calculating a gradient and an outer product of two vectors of length six. This requires a total of $T_i + p(T_g + 48T_a)$ operations.

The total computation for five iterations can now be written as:

$$NM(T_g + T_a) + 5(T_i + p(T_g + T_s + 59T_a)). \tag{5}$$

This expression can be reduced further by specifying the complexity of each operation in terms of one basic unit. A reasonable set of simplifications is to use bilinear interpolation for image resampling and a first-order difference for gradient calculations. A non-optimized bilinear interpolation operation requiring 15 arithmetic operations gives values of $T_s = 15T_a$ and $T_g = 16T_a$. Therefore a simplified expression for the floating point operations required to estimate motion parameters at a rate of 30 frames per second is given by:

$$150T_i + (510NM + 13500p)T_a. \tag{6}$$

Clearly the value of p should be as small as possible to reduce complexity. In our simulations ,the image size is 256×256 and less than 5% of the pixels are used to compute the motion parameters. Neglecting the time for matrix inversion, this leads to a computational burden of approximately 78 MFLOPS to process 30 frames per second.

5 SPATIO-TEMPORAL FILTERING

Once the global motion parameters have been calculated, this information can be used to reduce the noise in the sequence by combining a motion compensated filter together with the spatial Wiener filter previously used. Suppose that the motion parameters have been estimated and stored for the previous N frames. If the motion parameters for frame i are represented by $(\overline{\alpha}_i, \overline{\beta}_i)$, then Equation (1) can be recursively applied to find a mapping between the current frame and each of the previous N frames. For example, with $N = 2$ we can write:

$$X[n_1, n_2, k] = X[n_1 + \overline{\alpha}_k^T \overline{\theta}(n_1, n_2), n_2 + \overline{\beta}_k^T \overline{\phi}(n_1, n_2), k - 1] \tag{7}$$

$$X[n_1, n_2, k] = X\left[n_1 + \overline{\alpha}_{k-1}^T \overline{\theta}(n_1 + \overline{\alpha}_k^T \overline{\theta}(n_1, n_2), n_2 + \overline{\beta}_k^T \overline{\phi}(n_1, n_2)), \tag{8}\right.$$

$$\left. n_2 + \overline{\beta}_{k-1}^T \overline{\phi}(n_1 + \overline{\alpha}_k^T \overline{\theta}(n_1, n_2), n_2 + \overline{\beta}_k^T \overline{\phi}(n_1, n_2)), k - 2\right].$$

Assuming the global motion model is accurate, the mapping functions provide a way to obtain N additional observations of each pixel. Simple averaging of all observations reduces the noise variance by a factor of $N + 1$. Further reductions in noise are obtained by applying a spatial filter similar to Equation (4). The resulting spatio-temporal filter can be viewed as a two-step process where the N previous frames are first warped to align objects present in each of the frames. These warped frames are then averaged and spatially filtered.

6 RESULTS

The spatio-temporal noise suppression scheme was tested on a simulated video sequence. The simulated sequence was generated using the Georgia Tech Signature (GTSIG) model together with simulated target trajectories. GTSIG calculates the thermal signatures of a target by solving a system of differential equations. A series of sequences was generated to test different lighting conditions and interceptor velocities. Figure 3a-b shows two original frames from a typical sequence. The field of debris is composed of 64 fragments where each fragment is the projection of a 3-D object experiencing 3-D motion onto the 2-D image plane. The degraded images observed

at the sensor are show in Figure 3c-d. These images are produced by the model in Figure 2 where $H(w_1, w_2)$ is a Gaussian blur filter and the noise is uncorrelated Gaussian noise. The resulting SNR for the observed sequence is 4.3 dB for frame 240.

Two approaches to noise suppression are shown in Figure 4. The top images show the result of applying the Wiener filter given in Equation (4) to each of the observed images. Likewise, the bottom images show the result produced by the spatio-temporal noise suppression algorithm where three frames are used in the filter. In terms of SNR, both filters increase the SNR by 3 dB. The ultimate goal, however, is detection and tracking of these objects. This goal requires the objects to be easily distinguished from the background with minimal degradation in shape. Subjectively, the objects are easier to distinguish from the background and and outlines are more accurate in the frames processed with the spatio-temporal filter. This difference is objectively demonstrated in Figure 5 by thresholding each of the filtered images to generate a binary object mask. The threshold was selected for each image to be as low as possible without introducing any spurious objects into the background. The lower images in Figure 5 clearly show that more objects are successfully identified. In addition, objects boundaries of large objects are more accurate.

7 CONCLUSIONS

We have presented a spatio-temporal noise suppression algorithm intended for use as a preprocessing step in ballistic missile tracking and detection systems. As part of the noise reduction algorithm we calculate global motion estimates that can serve as initial estimates of target trajectories in subsequent processing stages. While the computation required to implement the algorithm can be reduced further by using a multi-resolution scheme, the current implementation yields good results with a computation burden of approximately 80 MFLOPS. Future research will apply the generalized motion estimation algorithm used in this work to the task of computing local motion information around targets of interest.

8 ACKNOWLEDGMENTS

9 REFERENCES

[1] J. M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Visual Communication and Image Representation*, vol. 4, pp. 348–365, December 1995.

[2] C. Kenney, G. A. Hewer, and W. Kuo, "Optic flow methods for tracking problems," *SPIE - Wavelet Applications*, vol. 2242, pp. 561–572, April 1994.

[3] J. W. Monaco and M. J. T. Smith, "Video coding using image warping within variable size blocks," in *ISCAS '96*, vol. 2, pp. 794–797, May 1996.

[4] W. Snyder and S. Rajala, "Image modeling the continuity assumption and tracking," in *Proceedings of the 5th International Conference on Pattern Recognition*, pp. 1111–1114, December 1980.

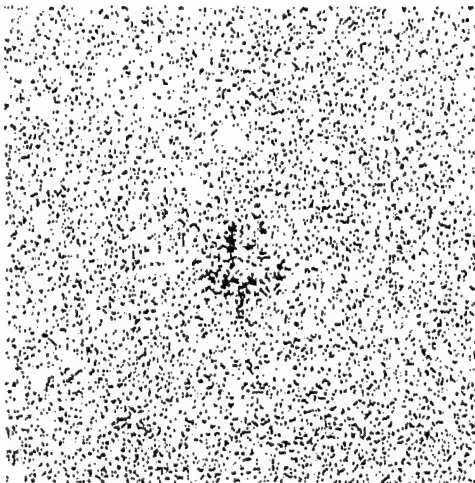[5] G. Wolberg, *Digital Image Warping*. IEEE Computer Society Press, 1990.
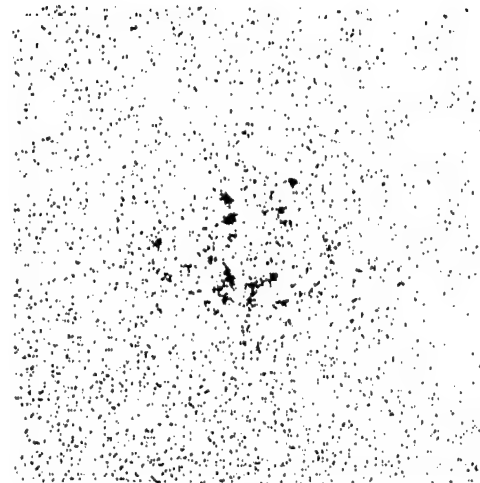
Original Frame 200   Original Frame 240

(a)   (b)
Observed Frame 200   Observed Frame 240
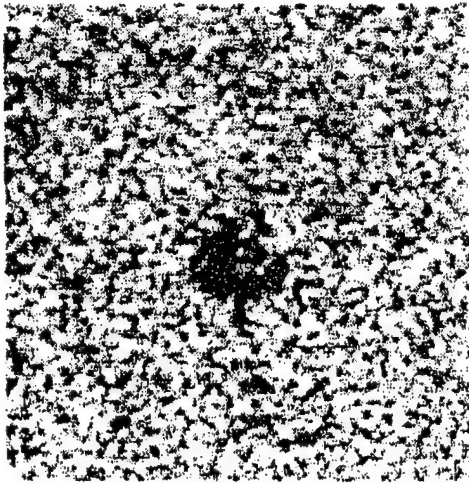

(c)   (d)

Figure 3: Selected frames from the simulated data sequence used for testing. The top row shows the original signal while the bottom row shows the observed signal that has been corrupted by a blurring distortion and noise.

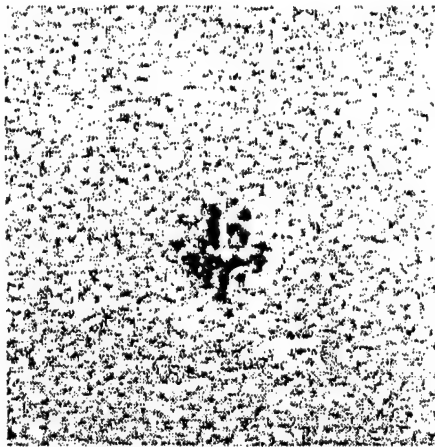Wiener filtered images

(a)　　　　　　　　　　　　　　(b)

Spatio-temporal filtered images

(c)　　　　　　　　　　　　　　(d)

Figure 4: Comparison of filtered frames. The left column compares results at frame 200 while the right column compares results at frame 240.

(a)

(b)

(c)

(d)

Figure 5: Comparison of threshold applied to filtered frames. The left column compares results at frame 200 while the right column compares results at frame 240.

Appendix D

Introduction

The objective of this project is the development of an effective algorithm for tracking a missile warhead when missile fragments are in the scene. The general approach taken is to study the problem within the framework of the Continuous Multidimensional Wavelet Transform (CWT) and to employ principles from multirate signal processing to achieve efficient implementations.

A substantial effort toward CWT tracking is in progress. This part of the project is described in Appendices A and B. Supporting the CWT tracking task is a motion estimation and modeling component, which we discuss next. Motion modeling may be used within the CWT or within any multiresolution framework. This is envisioned to be a critical part of the end product.

Motion estimation refers to the task of calculating the movement of objects through a series of images. This task is important in a variety of applications such as video coding, computer vision, and tracking. In general, two main approaches are used for the task of motion estimation. Block-based approaches assume a translational motion holds for every pixel in a given region. This assumption leads to straightforward computation of motion parameters with a search technique, but the limitation on translational motion requires small block sizes and is usually only accurate for short periods of time. In the second approach, pel-recursive techniques allow a unique motion vector for each pixel in an image. In order to make this approach tractable, neighborhood constraints are imposed on the motion field. These constraints require the motion field to be spatially smooth. Pel-recursive techniques can represent the true motion field much more accurately and any type of motion is possible. Unfortunately, this flexibility comes at the cost of high computation and reduced robustness to noise. We use a combination of these two approaches based on an explicit model of the motion field. Similar to block-based approaches, we assume the movement of pixels within a patch is governed by the same model. At the same time, the model gives a unique motion vector for every pixel in the patch like pel-recursive based techniques. By increasing the model order, more complicated motion can be represented. The advantages of this approach include better representation of true motion fields relative to block-based approaches, fewer computations than pel-recursive techniques, and a compact representation of the motion field with model coefficients.

General Algorithm

The following derivation uses arbitrary basis functions to show how a motion model of arbitrary complexity can be used. The model relating the current image at time index t to the previous image is given by

$$X(n_1, n_2, t) = X\left(n_1 + \sum_{i=1}^{p} \alpha_i \theta_i(n_1, n_2), \; n_2 + \sum_{i=1}^{q} \beta_i \phi_i(n_1, n_2), \; t-1\right) \quad (1)$$

where $X(n_1, n_2, t)$ represents the video signal luminance and $\theta_i(n_1, n_2)$ and $\phi_i(n_1, n_2)$ are basis functions used to represent the motion field. The goal of the motion estimation algorithm is to find optimal values for the coefficients α_i and β_i. In order to solve equation (1) we make two simplifications. First, the model coefficients are assumed to remain constant in an arbitrary 2D region denoted $\mathcal{S}$. Second, the expression on the right-hand side of equation (1) is approximated by a first-order Taylor series expansion as follows:

$$X\left(n_1 + \sum_{i=1}^{p} \alpha_i \theta_i(n_1, n_2), \; n_2 + \sum_{i=1}^{q} \beta_i \phi_i(n_1, n_2)\right) \approx$$

$$X(n_1, n_2) + \frac{\partial X(n_1, n_2)}{\partial n_1} \sum_{i=1}^{p} \alpha_i \theta_i(n_1, n_2) + \frac{\partial X(n_1, n_2)}{\partial n_2} \sum_{i=1}^{q} \beta_i \phi_i(n_1, n_2). \quad (2)$$

Combining these simplifications with the motion model leads to the following error function:

$$\sum_{[n_1, n_2] \in \mathcal{S}} \sum W(n_1, n_2) \left(X(n_1, n_2, t) - X(n_1, n_2, t-1) - \frac{\partial X(n_1, n_2)}{\partial n_1} \sum_{i=1}^{p} \alpha_i \theta_i(n_1, n_2) \right.$$

$$\left. - \frac{\partial X(n_1, n_2)}{\partial n_2} \sum_{i=1}^{q} \beta_i \phi_i(n_1, n_2) \right)^2. \quad (3)$$

where $W(n_1, n_2)$ is a weighting function. The minimum of the error function is found by forcing the derivative with respect to each α_i and β_i to zero. As long as a solution exists the minimum is found a $(p+q) \times (p+q)$ matrix inverse. The accuracy of the estimate is increased by changing the location used by the Taylor series approximation in equation (2) from (n_1, n_2) to

$$(n_1 + \sum_{i=1}^{p} \alpha_i \theta_i(n_1, n_2), n_2 + \sum_{i=1}^{q} \beta_i \phi_i(n_1, n_2)).$$

The motion parameters α_i and β_i are iteratively updated by forcing the derivative of the error function in expression 3 to zero and changing the location of the Taylor series expansion. This process can be repeated until either an error criterion or the maximum number of iterations is met. For the case of an affine model, optimal results are almost always reached in 5 or fewer iterations. An important consideration in this algorithm is the choice of initial starting points. In general the assumption of linear variation in intensity is only valid in a small

8

neighborhood around each pixel. Consequently, the update term used in each iteration of the algorithm must be fairly small. This fact implies that the starting point must be reasonable good. In typical video sequences this restriction does not present a serious problem because the temporal sampling rate is high enough to allow using a trivial starting point where all parameters are zero. In a tracking setting a better choice is to assume the motion parameters change slowly over time so that previous model parameters may be used for the initial estimate at the current time. This prediction based approach was used in the initial simulations.

Simulations

The general algorithm was used to track objects in test sequences composed of either 3 or 64 objects. The tracking algorithm consists of three parts.

1. Registration: Small changes in position of the camera result in relatively large translation displacements of target locations. Registration approximately compensates for these translation displacements. This step is important for future processing because it provides a fixed origin for the motion models.

   Registration is achieved with a simple algorithm that finds the centroid of all objects in the image. Once the centroid is located the entire image is shifted so that the centroid moves to the center of the image.

2. Global Motion Estimation: Part of the apparent motion in the sequence is caused by the decreasing distance between the camera and the targets. We chose to use an affine model for this component of target motion. For this case $p = q = 3$ and the basis functions are given by:

$$\theta_1(n_1, n_2) = \phi_1(n_1, n_2) = 1$$
$$\theta_2(n_1, n_2) = \phi_2(n_1, n_2) = n_1$$
$$\theta_3(n_1, n_2) = \phi_3(n_1, n_2) = n_2.$$

   This component of the motion applies to all the targets equally, and the entire image is used to estimate the parameters. Also, since the images are registered in step 1, we can use the previously estimated global motion parameters as an initial estimate of the motion parameters relating the current frame to the previous frame. Good initial estimates reduce computation and produce more accurate motion estimates.

3. Local Motion Estimation: Finally, a small correction in motion parameters is made for any targets of interest. As in step 2 an affine model is used to approximate the motion. Target selection is done by drawing a box around the desired object early in the sequence. Local motion parameters

9

are assumed to be valid within the box and the boundaries of the box are updated in each frame with the global motion parameters.

Future Work

The simulation results show the algorithm can reliably track the objects in the distortion free simulated sequence. Further study is needed to determine if the algorithm is appropriate in realistic scenarios where images are degraded by noise and blurring. Since the algorithm depends on reliable image gradients near edges, noise will be the most difficult problem. To address this problem, a robust gradient estimation algorithm must be used.

In addition to performance with noisy inputs, an important consideration in real time operation is the computational complexity. Simplifications to the algorithm that reduce complexity will be explored and the total computational cost in terms of floating point operations per second will be evaluated.
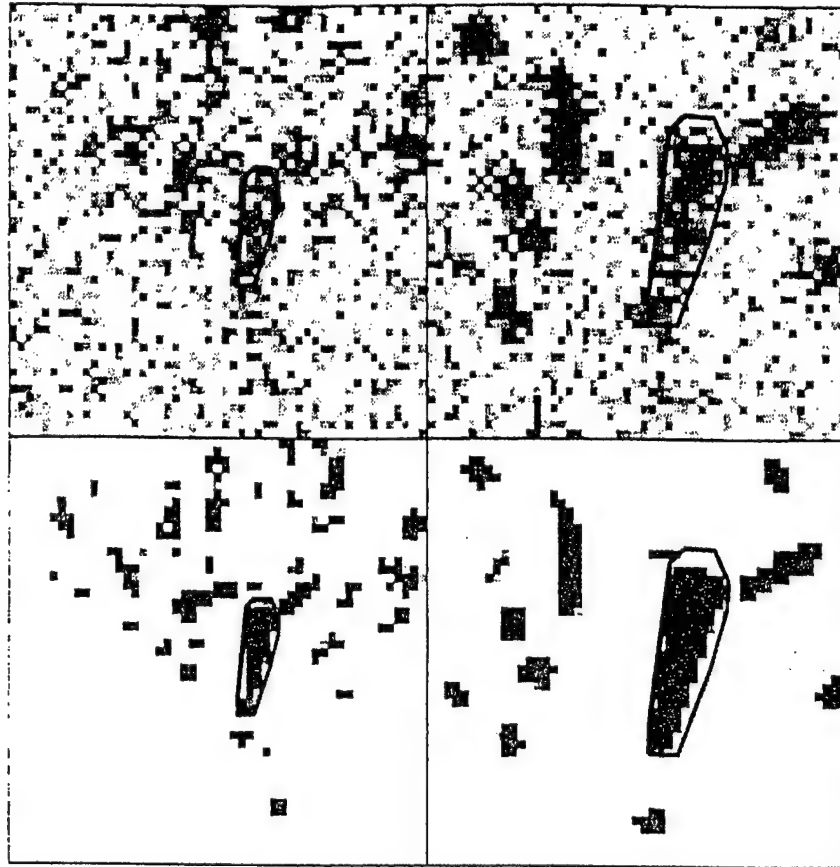
Figure   Summary of tracking results.

Appendix E

Linear-Prediction-Based Motion Estimation

As part of the Ph.D. thesis of Robbie Armitano, a computationally efficient block-matching motion estimation algorithm was developed using linear prediciton. The idea is that it narrows the search region using a linear prediction of the motion vectors. It has been found that a reliable estimate of motion cannot only reduce the computational complexity of motion estimation, but can also be used to (1) decrease the amount of side information required to transmit motion vectors, (2) recover lost motion vectors, and (3) improve the performance of motion estimation in noisy environments. We examined the use of linear prediction in the context of motion analysis for target tracking. In particular, we were interesting in how well the noise resiliant properties of the technique working in the tracking problem.

Linear prediction of motion vectors is possible because of the high correlation between motion vectors of neighboring blocks. Because of the high motion-vector correlation, one can use a linear combination of previously computed motion vectors to obtain an initial guess (a seed) for the estimation of the current motion vector, $\hat{v}_b(t)$, as seen in Eq. (4). The biasing of the search origin reduces the ambiguity in the motion-estimation function, minimizing the search area. The estimate is based on motion vectors in a neighborhood around the current motion vector that were previously computed in past, current, or future frames. Motion vectors that are highly correlated are used as a 3-D region of support, Ψ. The equation for the predicted motion vector, $\hat{v}_b(t)$, in frame t with a pth order prediction filter is given as

$$\hat{v}_b(t) = \sum_{k=1}^{p} \alpha_k v_{b_k}(t_k), \tag{4}$$

where b is the current block, t is the current frame number, $\{b_k, t_k\} \in \Psi$, and the α_k are the predictor coefficients.

The use of linear prediction to estimate a seed (an initial starting point) for the motion-vector search is advantageous because it reduces the search time by reducing the number of redundant searches. It imposes continuity from block to block and frame to frame. This forces the motion vector field to better represents the underlying optical flow and provides resiliance in noisy environments.

The linear prediction algorithm performs well in noisy-free environments as expected. This is illustrated by the example shown in Figure 2. It shows a single frame taken from our three-object test sequence. The motion vectors obtained are correct.

It can handle noise well at a level of 10-20 percent, but breaks down beyond this range. Figure 2 shows the resulting motion vectors for our tracking test sequence under realistic noise conditions. As can be seen, the algorithm derails.
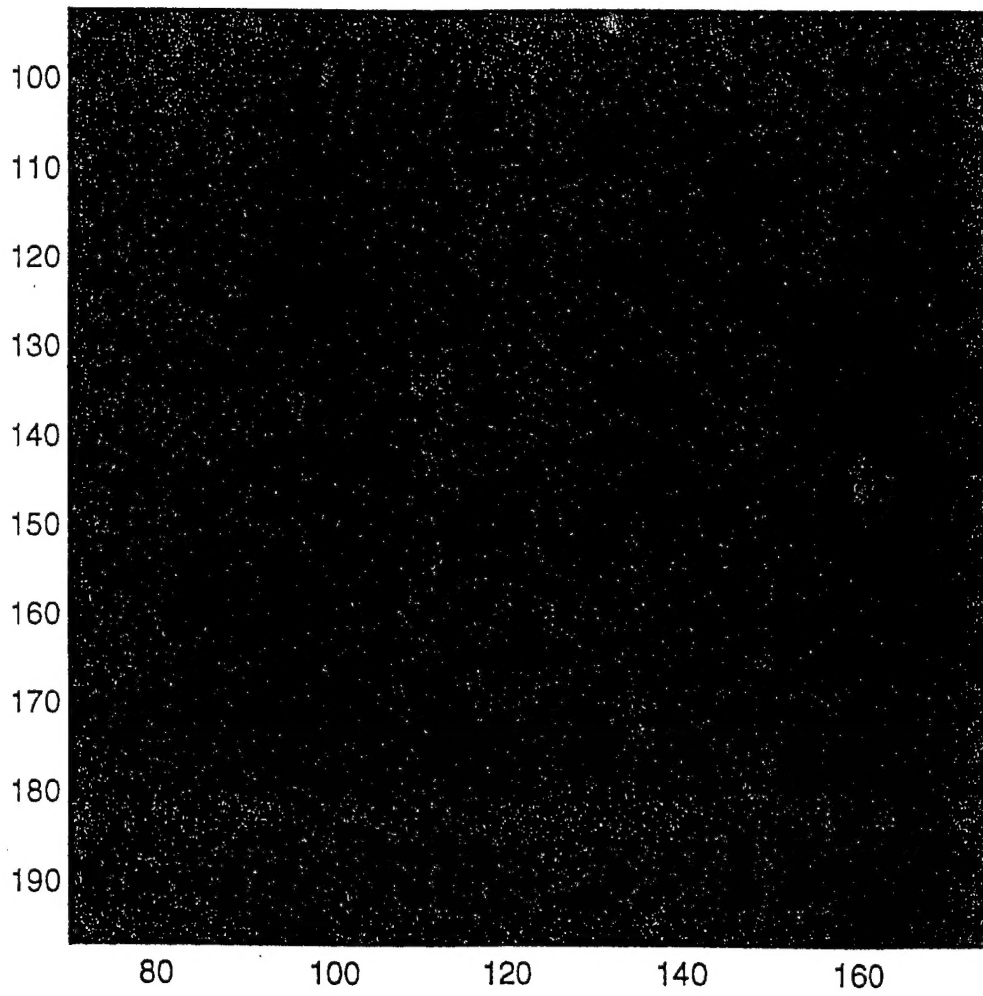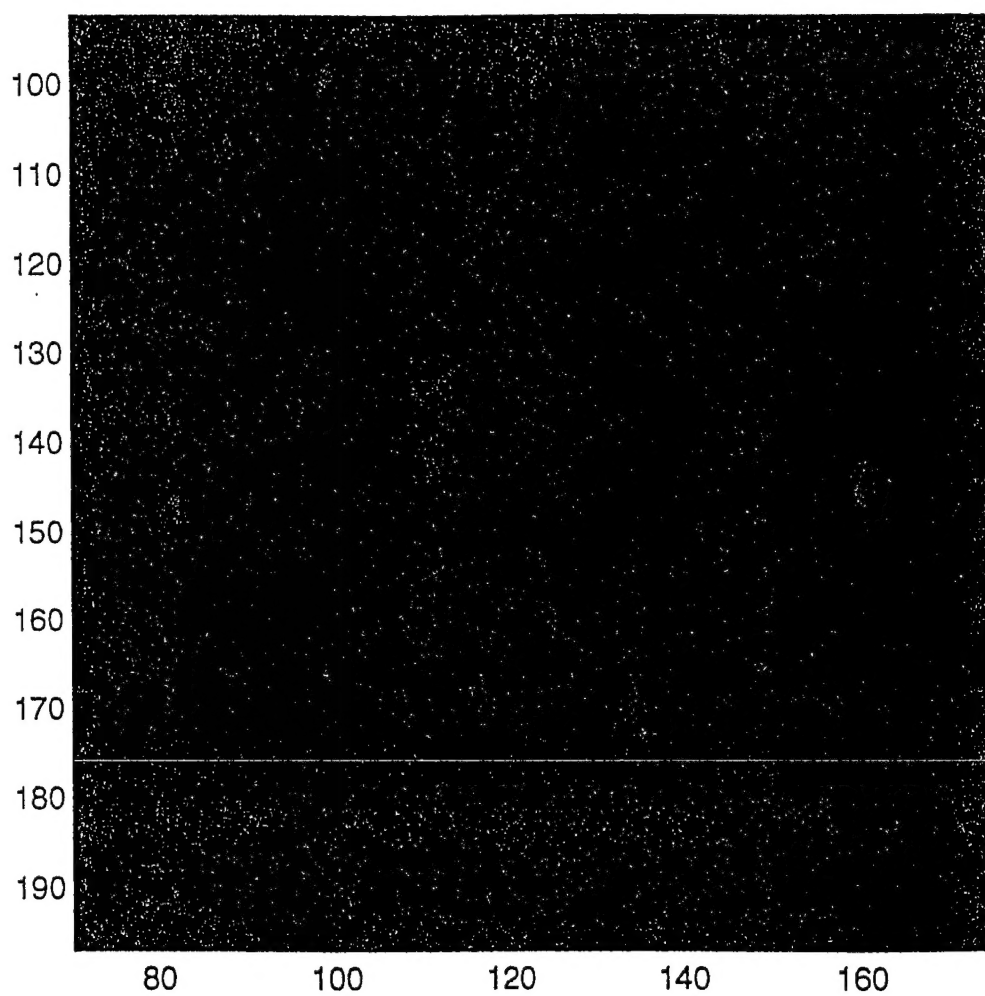
t1: Frame 217 SD: 2

Figure 2: Motion vectors obtained for noisy case.

Figure 2: Motion vectors obtained for noisy case.